

CIS-ACTING ELEMENTS CONTROLLING THE EXPRESSION OF THE HUMAN GLI3 GENE

Dissertation
zur
Erlangung des Doktorgrades
Der Naturwissenschaften
(Dr. rer. nat.)

dem
Fachbereich Biologie
der Philipps-Universität Marburg, Germany

vorgelegt von

Zissis Paparidis
aus Leverkusen

Marburg/Lahn Oktober 2005

CIS-ACTING ELEMENTS CONTROLLING THE EXPRESSION OF THE HUMAN GLI3 GENE

Dissertation
zur
Erlangung des Doktorgrades
Der Naturwissenschaften
(Dr. rer. nat.)

dem
Fachbereich Biologie
der Philipps-Universität Marburg, Germany

vorgelegt von

Zissis Paparidis
aus Leverkusen

Marburg/Lahn Oktober 2005

Vom Fachbereich Biologie

der Philipps-Universität Marburg als Dissertation am _____

_____ angenommen.

Erstgutachter _____

Zweitgutachter _____

Tag der mündlichen Prüfung am _____

CONTENTS

1	<u>SUMMARY</u>	1
2	<u>ZUSAMMENFASSUNG</u>	3
3	<u>INTRODUCTION</u>	5
3.1	The hedgehog (hh) signaling cascade	5
3.2	The human <i>GLI</i> family	10
3.3	Pathogenic effects of GLI mutations	11
3.4	Transcriptional regulation of genes of the Hh signaling cascade	13
3.5	Transcriptional regulation of co expressed genes	14
3.6	Highly conserved regions outside of coding sequences	15
3.7	Evolutionary sequence comparison facilitates the detection of long-range regulatory elements	15
3.8	Transcriptional regulation of paralogous genes	16
3.9	Objectives of the study	18
4	<u>MATERIALS AND METHODS</u>	20
4.1	MATERIALS	20
4.1.1	Devices and accessories	20
4.1.2	Chemicals	22
4.1.3	Buffers and standard solutions	23
4.1.3.1	DNA Isolation	23
4.1.3.2	Gel Electrophoresis	24
4.1.3.3	Bacterial culture	24
4.1.3.4	Cell culture	24
4.1.4	Enzymes	25
4.1.5	Loading dye	25
4.1.6	Primers	26
4.1.6.1	Primers for SSCA	26
4.1.6.2	Primers for cloning	28
4.1.6.3	Primers for mutagenesis	29
4.1.6.4	Primers for RACE PCR	30
4.1.6.5	Primers for primer extension	30
4.1.6.6	Primers for sequencing	30
4.1.7	DNA size standards	30
4.1.8	Reaction kits	31
4.1.9	Plasmids used in this work	31
4.1.10	Organisms	37
4.1.10.1	Chemically competent bacteria	37
4.1.10.2	Cell lines	37
4.1.11	Data banks and online software	38
4.2	METHODS	38
4.2.1	Single strand conformation analysis (SSCA)	38

4.2.1.1	PCR amplification of fragments of the GLI3 coding region	38
4.2.1.2	Non-denaturing polyacrylamide gel for mutation analysis	39
4.2.1.3	Silver staining of the native polyacrylamide gels	39
4.2.1.4	Isolation of DNA from polyacrylamide gels	39
4.2.1.5	Sequencing PCR	39
4.2.1.6	Sequencing PCR precipitation	40
4.2.1.7	Sequencing electrophoresis and data analysis	40
4.2.2	Reverse transcription PCR (RT-PCR)	40
4.2.3	Polymerase chain reaction – PCR	40
4.2.4	Agarose gel electrophoresis of DNA	41
4.2.5	5' RACE PCR	41
4.2.6	Plasmid isolation using QIAGEN QIAprep spin miniprep kit	42
4.2.7	Medium-scale preparation of plasmid DNA (Qiagen Midi- Kit)	42
4.2.8	Large-scale preparation of plasmid DNA (Qiagen Maxi- Kit)	43
4.2.9	RNEASY Mini Protocol for total RNA isolation from cells	43
4.2.10	QIAquick Gel Extraction method	44
4.2.11	Restriction enzyme digestion of DNA	44
4.2.12	Klenow treatment	44
4.2.13	CIP treatment of digested vector	45
4.2.14	Ligation (Invitrogen)	45
4.2.15	TOPO TA Cloning	45
4.2.16	Chemical transformation of TOP 10 E.coli competent cells	45
4.2.17	Liquid culture	46
4.2.18	Small-scale isolation of plasmid DNA	46
4.2.19	Plasmid DNA sequencing	46
4.2.20	Cell techniques	47
4.2.20.1	Culture of eukaryotic cells	47
4.2.20.2	Freezing cultured eukaryotic cells	47
4.2.20.3	Splitting the eukaryotic cell cultures	47
4.2.20.4	Maintenance of human cell lines	47
4.2.20.5	Cell counting	48
4.2.20.6	Transient transfection of adherent cells	48
4.2.20.7	Protocol for dual luciferase assay on a 12 well plate	49
4.2.21	Primer extension assay	49
4.2.21.1	Primer extension reaction	50
4.2.21.2	Precipitation of primer extension product	50
4.2.22	Transgenic mice	50
4.2.22.1	DNA isolation and precipitation prior to injection	51
4.2.22.2	Mice injection	51
4.2.22.3	Analysis of embryos	51
4.2.22.4	DNA isolation from mice	52
5	<u>RESULTS</u>	54
5.1	Search for mutations in individuals with putative GLI3 morphopathies	54
5.2	Experimental control of the published human GLI2 cDNA sequence	57
5.2.1	Human GLI2 5'RACE PCR	57
5.3	Identification and functional analysis of the GLI3 promoter	60
5.3.1	Determination of transcription start site of human GLI3	60

5.3.2	Experimental determination of the transcription start site with the primer extension assay	62
5.3.3	Cloning of promoter constructs	63
5.3.4	Analysis of the capability of the promoter constructs to drive reporter gene expression in cell culture	64
5.3.5	Search for critical transcription factor binding sites in the <i>GLI3</i> minimal promoter	66
5.4	Identification of conserved elements downstream of <i>GLI3</i> promoter	70
5.5	Highly conserved intronic regions of <i>GLI3</i>	73
5.5.1	Transcriptionally important sites within the Hom2 region	79
5.5.2	Reporter gene regulation by Hom2 mutants in H661 cells	80
5.5.3	Reporter gene regulation by Hom2 in different cell lines	81
5.5.4	NFATP as a candidate transcription factor to interact with Hom2	82
5.5.5	Search of a transcriptional start by primer extension analysis on Hom2	83
5.5.6	Deletion mapping of functionally important segments of Hom3	85
5.6	The potential of the homology regions Hom2, Hom3, and Hom4 to drive reporter gene expression in transgenic mice	87
5.6.1	Stained embryos	89
5.6.1.1	Embryo 13 - 9 days	89
5.6.1.2	Embryo 11 - 11 days	90
5.6.1.3	Embryo 30 - 11 days	92
5.6.1.4	Embryo 32 - 11 days	93
6	<u>DISCUSSION</u>	96
6.1	GLI3 morphopathies without mutations in transcribed sequences are candidates for defects in regulatory sequences	97
6.2	Human GLI2 is lacking DNA sequence homology with GLI3 in the 5' region and the promoter	100
6.3	Analysis of the GLI3 minimal promoter	104
6.4	A region in the 5' part of intron 1 enhances the action of the GLI3 minimal promoter	108
6.5	Effect of highly conserved intronic regions on the expression of a luciferase reporter gene	110
6.6	Hom2 but not Hom3 and Hom4 direct tissue specific expression of a reporter gene in transgenic mice	116
6.7	The reporter gene expression pattern appears to recapitulate GLI3 activity	117
6.8	Hom2 is a conserved non-genic sequence element acting as transcriptional enhancer	120
7	<u>OUTLOOK</u>	125
8	<u>ABBREVIATIONS</u>	127
9	<u>REFERENCES</u>	129
10	<u>PUBLICATIONS</u>	147
10.1	Original work	147

10.2	Posters	147
10.3	Seminars	147
11	<u>ACKNOWLEDGEMENTS</u>	148
12	<u>DECLARATION</u>	150
13	<u>CURRICULUM VITAE</u>	151

1 SUMMARY

Limb development is a complex mechanism involving many molecular factors and signaling pathways, which orchestrate during embryogenesis the transformation of the limb bud to the complete extremities. Posterior-anterior patterning is directed by the sonic hedgehog signal molecule, which acts upon transcriptional activity of target genes via the GLI transcription factors. Time, location, and amount of the transcription of *GLI*-genes are of critical importance. Mutation affecting the availability of the appropriate amounts of these factors in limb bud cells as well as mutations impairing their function can result in developmental defects and tumorigenesis.

To contribute to the detection and functional characterization of cis-acting regulatory elements for *GLI3* and their potential relevance for pathogenity three questions were addressed in this thesis.

- How is the expression of human *GLI3* regulated?
- Do *GLI2* and *GLI3* share similar regulatory elements?
- Are mutations in *GLI3* regulatory elements involved in pathogenity?

Towards this end, 24 patients with limb defects classified as potential *GLI3* morphopathies were screened for mutations. 20 cases, which cannot be attributed to a mutation in coding sequences of the gene, are candidates to be searched for mutations in cis-regulatory elements.

The transcriptional control of *GLI3* gene expression involves promoter as well as cis-acting sequences, such as enhancers. A minimal promoter was defined and tested functionally. Two initiator sites were identified by using templates from placenta and skeletal muscle. By mutagenesis, sequence elements involved in control of *GLI3* expression were identified.

Functional studies in transgenic mice suggest that *GLI2* and *GLI3* might have greatly overlapping expression domains. The 5' transcribed region of human *GLI2* was extended by about 1 kb of noncoding DNA, however, sequence comparison of human and murine *GLI2* or *GLI3* did not detect homologies of regulatory elements.

Evolutionary genomic sequence comparisons were applied to guide the search for highly conserved non-coding elements, which might represent cis-regulatory elements for *GLI3*. Three potential enhancer elements were tested for their regulatory capacity

on reporter genes with foreign as well as the original *GLI3* minimal promoter in transiently transfected cell cultures and in transgenic mice. Mutagenesis followed by tests for retention of their regulatory capacity in cellular reporter gene assays pinpointed particularly critical sites. Transcription factors that could be involved in *GLI3* regulation, such as NFATp, await independent confirmation. In transgenic mouse embryos it was determined, that one of the potential enhancer elements directs an expression pattern, which follows part of the time course and the spatial pattern of the endogenous mouse *GLI3*, in particular the brain, mandibles, nostrils and heart.

The results obtained contribute to our understanding of the spatial and temporal control of the expression of *GLI3*, a key factor of the hedgehog signaling cascade and provide insight into the potential role of highly conserved non-coding sequence elements in the human genome.

2 ZUSAMMENFASSUNG

Die Entwicklung der Gliedmaßen ist ein komplexer Vorgang, der zahlreiche molekulare Faktoren und Signalwege involviert. Diese dirigieren während der Embryogenese die Umwandlung der Gliedmaßenknospe zu den fertigen Extremitäten. Die antero-posteriore Musterbildung wird vom Signalmolekül „sonic hedgehog“ gesteuert, das über die GLI-Transkriptionsfaktoren die Transkriptionsaktivität von Zielgenen reguliert. Zeit, Lokalisierung und Menge der *GLI*-Transkription sind von kritischer Bedeutung. Mutationen, die das Vorhandensein der notwendigen Menge dieser Faktoren in den Zellen beeinflussen oder ihre Funktion stören, können in den Gliedmaßenknospen zu Wachstumsdefekten und andernorts zur Krebsentstehung führen.

Um einen Beitrag zum Nachweis und zur funktionellen Charakterisierung von cis-regulatorischen Elementen von *GLI3* und deren potentieller Bedeutung für die Pathogenität zu liefern, wurden in dieser Dissertation drei Fragen bearbeitet:

- Wie wird die Expression von humanem *GLI3* reguliert?
- Besitzen *GLI2* und *GLI3* ähnliche regulatorische Elemente?
- Sind Mutationen in regulatorischen Elementen von *GLI3* an der Pathogenität involviert?

Um letzteres vorzubereiten wurden 24 Patienten mit Gliedmaßendefekten, die als potentielle *GLI3*-Morphopathien klassifiziert waren, auf Punktmutationen hin untersucht. 20 Fälle, bei denen keine Mutation in der kodierenden Sequenz gefunden werden konnte, sind Kandidaten für eine Mutationssuche in cis-regulatorischen Elementen.

An der Kontrolle der Transkription des *GLI3*-Gens sind die Promoterregion und andere cis-aktive Sequenzen, wie Enhancer, beteiligt. Ein minimaler Promotor wurde definiert und funktionell getestet. Zwei Initiationsstellen der Transkription wurden mit cDNA aus Plazenta und Skelettmuskeln identifiziert. Durch Mutagenese wurden Sequenzelemente, die an der Kontrolle der *GLI3*-Expression beteiligt sind, eingegrenzt.

Funktionelle Studien an transgenen Mäusen hatten gezeigt, dass die Expressionsdomänen von *GLI2* und *GLI3* in großen Bereichen überlappen. Die transkribierte Region von *GLI2* wurde im 5'-Bereich gegenüber der bekannten cDNA

um 1 kb nicht kodierender DNA ausgedehnt. Trotzdem konnten Sequenzvergleiche zwischen Mensch und Maus keine Homologie zwischen regulatorischen Elementen von *GLI2* oder *GLI3* entdecken.

Die genomische Sequenz von *GLI3* wurde für evolutionäre Vergleiche benutzt, um nach hochkonservierten, nicht-kodierenden Elementen zu suchen, die eventuell cis-regulatorische Elemente repräsentieren. Drei so entdeckte potentielle Enhancerelemente wurden in transient transfizierten Zellen daraufhin untersucht, ob sie die Transkription eines Reportergens unter der Kontrolle des minimalen *GLI3*-Promoters steuern konnten. Durch Mutagenese vorhergesagter Bindungssequenzen für Transkriptionsfaktoren und nachfolgende Analyse der regulatorischen Kapazität der Elemente in zellulären Reportergentests wurden funktionell besonders bedeutsame Stellen definiert. An diese Bereiche bindende Transkriptionsfaktoren, die in die Regulation von *GLI3* involviert sein könnten, wie NFATp, müssen noch bestätigt werden. Die drei Elemente wurden ebenfalls auf ihre Fähigkeit hin untersucht, in transgenen Mäusen ein Reportergen zu aktivieren. In transgenen Mausembryonen konnte beobachtet werden, dass eines der potentiellen Enhancerelemente ein Expressionsmuster steuerte, das zeitlich und räumlich einem Teil des endogenen Maus-*GLI3*-Musters entsprach, vor allem im Gehirn, den Mandibeln, im Nasenbereich und im Herzen.

Die erzielten Ergebnisse tragen zum Verständnis der räumlichen und zeitlichen Expressionskontrolle von *GLI3* bei, dem Schlüsselfaktor der „sonic hedgehog“-Signalkaskade, und verschaffen einen Einblick in die potentielle Rolle hochkonservierter, nicht-kodierender Sequenzelemente im menschlichen Genom.

3 INTRODUCTION

Limb development is a complex mechanism involving many molecular factors and signaling pathways, which orchestrate during embryogenesis the transformation of the limb bud to the complete extremities. The same or homologous signaling molecules and mechanisms are used to regulate development of many organs and tissues of the body. There, they are not easily accessible for genetic studies in humans since mutations leading to serious malformation syndromes generally do not segregate in large pedigrees. Limb defects, however, often do not seriously affect viability in humans and, consequently, are observed in extended families. In addition, limbs are easily accessible in model animals for analysis and experimental manipulation. Therefore, limbs are a preferred model for developmental biology, and the basic mechanisms of development unveiled for limbs in many cases apply for development of organisms in general.

3.1 The hedgehog (hh) signaling cascade

Hedgehog (Hh) signalling is important for the development of the limb and many other tissues and organs, and damage to components of the Hh signal-transduction pathway can lead to birth defects and cancer in man. The significance of this pathway for the development is shown by the fact that it is represented in many different organisms of the animal kingdom, both vertebrates and invertebrates.

The hedgehog pathway has been originally elucidated for the model organism *D. melanogaster* [reviewed by (Hooper and Scott 2005)]. Whereas the basic Hh-regulatory mechanism has been conserved between fly and vertebrates, some important differences emerge: One crucial distinction is that single genes in *D. melanogaster* are represented by gene families in mammalian species. For example, there is only the one Hh in *Drosophila*, whereas in mammals there exist three different hedgehog paralogs, sonic, Indian, and desert hedgehog, with different patterns of expression. *Sonic hedgehog (Shh)* is expressed in the ventral neural tube and its lateral somites. It is involved in the establishment of the left-right asymmetry, the polarization during the limb development, and it functions as a morphogen for the development of the eyes, lungs, hair, organs of the digestive system, and the nervous system (Krauss et al. 1993; Riddle et al. 1993; Fan et al. 1995; Roelink et al. 1995;

Marti and Bovolenta 2002; Salinas 2003; Donner and Maas 2004; Ishibashi et al. 2005). *Indian hedgehog (Ihh)* is expressed in the embryonic lung, in the developing cartilage, and in the adult kidney and liver (Marigo et al. 1995; Iwamoto et al. 1999). It coordinates the proliferation and the maturation of the chondrocytes during the development of the cartilage skeleton. Ihh regulates partly the same target genes as Shh, such as *Hoxd11*, *Hoxd13* and *Bmp2*, but also the parathyroid hormone-related peptide in the perichondrium cells (PTHrp) (Bitgood and McMahon 1995; Vortkamp et al. 1996; Iwamoto et al. 1999; St-Jacques et al. 1999). Finally, *Desert hedgehog (Dhh)* plays a role in the spermatogenesis (Bitgood et al. 1996).

The patched (*ptc*) receptor of the HH signaling molecule of *D. melanogaster* is represented by two *ptc* receptors in mammals, *Ptc1*, which is expressed adjacent to the hedgehog genes, and *Ptc2*, which is less well studied but appears to be not so widely expressed. Whereas the pathogenic effects of PTC1 deficiency has been well characterized, few phenotypes could be attributed to the malfunction of PTC2 (Rahnama et al. 2004).

Finally, the Drosophila transcription factor Cubitus interruptus (Ci) that translates HH signaling into regulation of gene expression, is represented in vertebrates by the *GLI* gene family, which consists of three members (*Gli1*, *Gli2*, and *Gli3*).

The current view of steps implicated in hedgehog signaling is presented in figure 3-1 for SHH

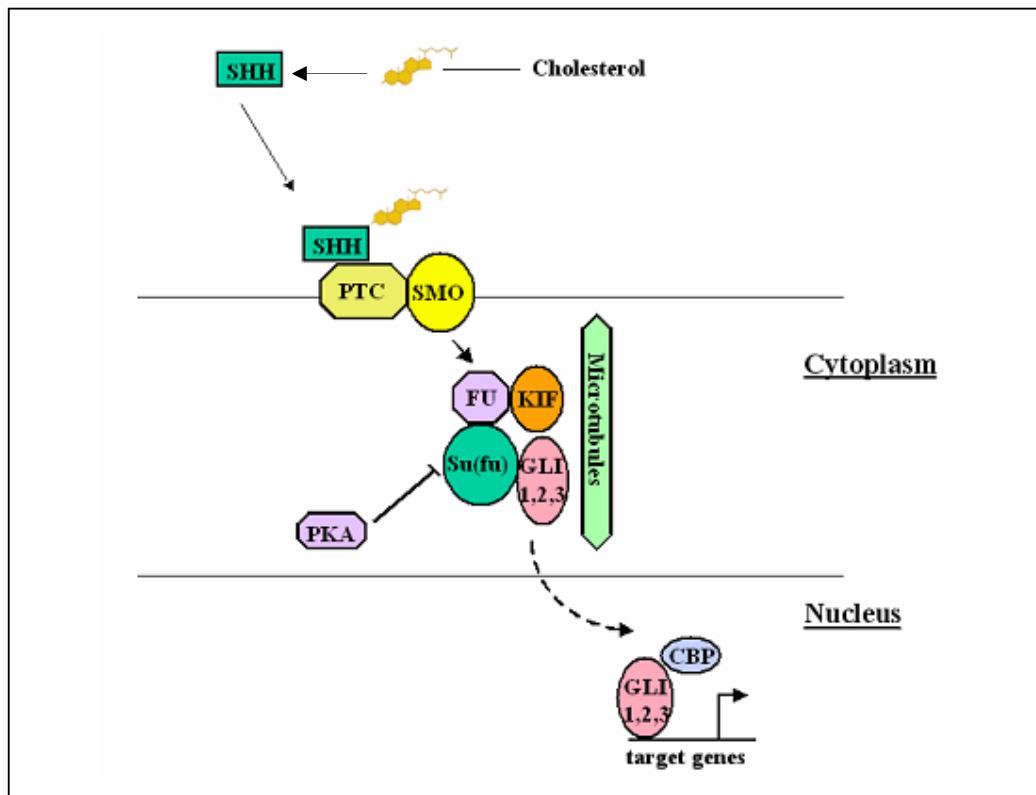


Figure 3-1: Shh signaling pathway

A selection of factors participating in the Shh signaling pathway. Autocatalytic cleavage and cholesterol binding is essential for the transport of Shh out of the signaling cell and to the target receptor as well as the action within the responding cell. Inactivation of the receptor Ptc after Shh binding releases and activates Smo. This event triggers a series of only partly understood reactions, and, eventually, to the transduction of the signal to the GLI proteins, which can act on their target genes

As a first step, Hh is autocatalytically processed and its N-terminal domain is covalently attached with a cholesterol and a palmitic acid molecule that plays an important role in its secretion and function in vertebrates (Pepinsky et al. 1998; McMahon 2000). Its receptor, the Ptc molecule, inhibits another transmembrane protein, the Smo molecule, in absence of Hh. This repression has the effect that Ci in *Drosophila* and Gli genes in vertebrates are maintained in a full length status, which can act as transcriptional activator. In the case of *Drosophila*, Smo is constitutively active but is inhibited by Ptc when Shh is not bound to it. The signaling by Hh causes the endocytosis of Ptc and the phosphorylation of Smo that transduces the signal into the cytoplasm (Chuang and Kornberg 2000; Deneff et al. 2000; Incardona et al. 2000; Kalderon 2000; McMahon 2000; Karpen et al. 2001). The factor Fu, which is a part of a multiprotein complex, bound to the microtubules receives the signal from Smo. This

complex consists of the proteins Fu, Su(fu), Cos-2 and Ci (Ingham 1998) but factors such as PkA, Slmb and Cbp also participate in this signal transduction pathway (Perrimon 1995; Aza-Blanc and Kornberg 1999). The Su(fu) is an antagonist for Fu and interacts directly with Cos-2 (Monnier et al. 1998). Cos-2 shows a high affinity to microtubules which is influenced by its phosphorylation status (Robbins et al. 1997; Sisson et al. 1997). Ci is the only DNA binding factor in this cascade and is associated with the microtubules through Cos-2 (Alexandre et al. 1996; Aza-Blanc and Kornberg 1999; Wang and Holmgren 1999). It is cleaved following the Hh signal. The N-terminal part of the molecule retaining the zinc-fingers and a regulatory domain enters the nucleus and represses the transcription of target genes. The phosphorylation status of Fu, Cos-2 and Ci is influenced by the PkA (Perrimon 1995; Chen et al. 1998) whereas Slmb plays a role in the ubiquitin mediated proteolysis through the proteasom and is connected with the activity of Ci (Ingham 1998; Jiang and Struhl 1998; Aza-Blanc and Kornberg 1999). Finally, Cbp is a general co-activator and repressor of many transcription factors (Giles et al. 1998).

In the absence of Hh, Ci represses its target genes whereas in the presence of Hh signaling it acts as an activator (Alexandre et al. 1996; Aza-Blanc et al. 1997). In vertebrates, this pathway is not as well elucidated as in *Drosophila*.

Ci and its vertebrate homologs, the GLI proteins, are regulated post-translationally in a similar way. Both of them are situated in the cytoplasm, and in the absence of Hh signaling they remain there bound to Sufu and Cos2 in *D. melanogaster* and to the Sufu and the KIF7 in vertebrates. Proteolytic cleavage of these proteins follows and a truncated form of them, missing the C' transactivating domain, represses the downstream target genes of this cascade (Kogerman et al. 1999; Methot and Basler 2000; Cheng and Bishop 2002; Merchant et al. 2004; Paces-Fessy et al. 2004; Tay et al. 2005). Pka processes Gli3 and perhaps Gli2 into a transcriptional repressor (Ruiz i Altaba 1999). Nevertheless, the GLI proteins can have different roles. Gli3 acts more as a repressor (Dai et al. 1999; Wang et al. 2000; Motoyama et al. 2003), Gli2 more as an activator and Gli1 only as an activator (Karlstrom et al. 2003). Hh signals activate the transcription of Gli1 through Gli2 and Gli3 (Bai et al. 2004)

Differences between the *D. melanogaster* and the vertebrates can also be found for the phosphorylation and regulation of the Smo molecule, as well as for the Cos2 factor. Orthologs of Cos2 in vertebrates are the KIF27 and KIF7 factors, members of the

kinesin family. KIF7 has been shown to bind GLI1 and to repress SHH signaling (Katoh and Katoh 2004; Tay et al. 2005). Another member of the KIF family, the KIF3a as well as the intraflagellar transport proteins (IFT88 and IFT172) play Cos-2 like roles in vertebrates, taking part in repression and activation of Hh signalling (Huangfu et al. 2003; Tay et al. 2005).

Some genes in the Hh signaling cascade of the vertebrates do not have any orthologs in *D. melanogaster*, some have orthologs with an unknown role and some have orthologs with completely different functions. “Missing In Metastasis” (MIM/BEG4), a actin-binding protein, functions as a Shh-responsive gene as well as a regulator of Gli transcription (Callahan et al. 2004). Two factors of the Hh pathway with positive regulatory activity that do not have orthologs in *D. melanogaster* are megalin and iguana. The first is a low-density lipoprotein-receptor protein that associates with SHH whereas the latter is a zinc-finger protein that participates in the nuclear localization of GLI1 (McCarthy et al. 2002; Sekimizu et al. 2004; Wolff et al. 2004). Two negative regulatory factors exclusively for the vertebrate pathway are FKBP8, an antagonist of SHH in the nervous system and SIL, a cytosolic protein with a function downstream of PTC (Izraeli et al. 2001; Bulgakov et al. 2004). One example of a factor that exists both in flies and vertebrates, but does not seem to affect the Hh pathway in *D. melanogaster* is Rab23. It is a vesicle transport protein and a negative regulator of the mouse Shh cascade (Eggenchwiler et al. 2001). Finally, an example of ortholog factors with different functions between the phyla is the Wnt inhibitory factor (WIF), which binds Wnt proteins and facilitates the Wnt signaling. Its vertebrate ortholog Shifted (Shf) is a secreted protein that binds Hh and takes part in its signaling. (Hsieh et al. 1999; Han et al. 2004; Glise et al. 2005; Gorfinkiel et al. 2005)

Some of the differences between organisms of the different phyla can be attributed to the fact that non-homologous factors have similar functions. HIP, the mammalian membrane glycoprotein Hh-interacting protein and Pxb of *D. melanogaster* probably have the same function despite the lack of sequence similarity. Both factors are transcriptionally regulated by Hh and both decrease the Hh responses, for example through direct binding to SHH in case of HIP (Chuang and McMahon 1999; Inaki et al. 2002).

3.2 The human *GLI* family

GLI proteins are the vertebrate transcription factors involved in the translation of hedgehog signals into gene regulation.

The first gene of this family to be described, *GLI1*, was identified as an over-expressed gene in a malignant glioma (Kinzler et al. 1988). Homology searches detected two more paralogs, *GLI2* and *GLI3*. All *GLI* genes encode proteins that contain five highly conserved tandem C₂-H₂ zinc fingers that are connected with a histidine-cysteine linker (Ruppert et al. 1988). The DNA-target sequence for the *GLI1* and *GLI3* proteins is a conserved GACCACCCA found in the promoters of the *GLI* target genes (Kinzler and Vogelstein 1990; Ruppert et al. 1990; Pavletich and Pabo 1993; Vortkamp et al. 1995), whereas *GLI2* has been reported to bind to a conserved GACCACCCA region (Tanimura et al. 1998).

In humans, all of the three genes have been mapped to separate chromosomal loci by FISH (Arheden et al. 1989; Ruppert et al. 1990; Kas et al. 1996; Matsumoto et al. 1996). Human *GLI1* resides in the 12q13.3 chromosomal region, spanning 13 kb of DNA sequence. It includes 12 exons with 3648 bp in total and is translated into a 1106 aa protein. It is regulated like its *Drosophila* homologue *ci* through a TATA-less promoter, lacks a CCAAT consensus sequence, has a high GC content, includes a CpG island, and contains several GC boxes (Schwartz et al. 1995; Liu et al. 1998). The human and mouse proteins are 85% conserved. Human *GLI2* resides in 2q14, but its genomic sequence and structure has not yet been unequivocally characterized. A 100% identity can be observed at the protein level between human and mouse *GLI2* proteins for the “GLI-homology domain 1” and for the DNA binding region (Hughes et al. 1997).

GLI3, finally, extends over a genomic region of 240 kb on chromosome 7p13. It consists of 15 exons with a 3 kb 3' untranslated fragment and gives rise to a 8,5 kb mRNA with a 5055 bp ORF, which corresponds to a protein of 1580 aa (Kinzler and Vogelstein 1990; Ruppert et al. 1990; Vortkamp et al. 1995; Wild et al. 1997). The *GLI3* gene expression is promoted by a 230 bp fragment within a CpG island, which resides adjacent to the 5' end of the 1st exon, whereas the translation start site is located in the 2nd exon (Jacobsen 1996). By comparing the *GLI* protein sequences it appears that *GLI3* and *GLI2* are functionally and structurally more similar to each

other than to GLI1. Sequence comparison between human GLI3 and mouse GLI2 showed that in addition to seven homologous domains they share with GLI1, they contain five more regions that cannot be found in GLI1. One resides before the 1st domain and the others after the domains 1, 4, 5 and 7. Also, in comparison to other factors of the GLI/Kruppel family, these two genes show the highest similarity (48%) on the protein level (Hughes et al. 1997). Due to the fact that both proteins interact with the CREB-binding protein (CBP) and have a similar expression patterns, it has been postulated that beside their unique roles they might possess also redundant functions. In the mouse, indeed redundancy has been shown with single and double GLI2/GLI3 knock-out animals (Mo et al. 1997; Hardcastle et al. 1998). Expression studies for the GLI genes have been conducted preferentially on mice (Buscher and Ruether 1998; Motoyama et al. 1998; Brewster et al. 2000; Schweitzer et al. 2000).

Gli3 is regarded as a mediator of Shh signaling and takes part in the development of the limbs, the lungs, and the brain. It is expressed in a wide variety of normal adult tissues, including lung, colon, spleen, placenta, testis, and myometrium. It has been shown that Gli3 plays a crucial role in the antero-posterior prepatternning of the forming limb, and it participates in the positioning of the limb through interactions with dHand and Tbx3 (Rallis et al. 2005) as well as the determination of the digit number and identity (Litingtung et al. 2002), the formation of scapular blade (Kuijper et al. 2005), the formation of skeletal muscles (McDermott et al. 2005), the patterning of the spinal cord (Warburton and Lee 1999; Jacob and Briscoe 2003), the lung (Warburton and Lee 1999), and the eye development (Tyurina et al. 2005).

3.3 Pathogenic effects of GLI mutations

Changes in the expression of GLI genes due to regulatory defects or mutations affecting the function of the proteins are associated with pathogenic phenotypes both in man and mouse.

GLI1 is overexpressed in glioblastomas, osteosarcomas, rhabdomyosarcomas, B cell lymphomas and basal cell carcinomas (Kinzler et al. 1987; Roberts et al. 1989; Dahmane et al. 1997; Werner et al. 1997; Ghali et al. 1999). Ectopic expression of human GLI1 in mice has been observed to cause developmental defects, failure to thrive and Hirschsprung-like dilatation of the gastrointestinal tract (Yang et al. 1997).

Defects in *GLI2* in mouse can cause basal cell carcinomas and skeletal disorders (Sasaki et al. 1999; Grachtchouk et al. 2000; Park et al. 2000). In humans, loss-of-function mutations can lead to pituitary anomalies and holoprosencephaly-like features (Roessler et al. 2003).

Point mutations, translocations and deletions throughout and flanking the *GLI3* gene can cause various autosomal dominant polysyndactyly syndromes such as the Greig cephalopolysyndactyly syndrome (GCPS), which is featured mainly by preaxial polydactyly, syndactyly, broad thumbs and toes, and facial deformities such as hypertelorism and frontal bossing (Vortkamp et al. 1991; Wild et al. 1997; Kalff-Suske et al. 1999). Frameshift and nonsense mutations can also lead to Pallister-Hall syndrome (PHS) which is characterized by hypothalamic hamartoma, central or postaxial polydactyly, syndactyly, imperforate anus, anteverted nares and other facial abnormalities, and, occasionally, associated HPE and malformations of the axial skeleton (Kang et al. 1997). Finally, nonsense and missense *GLI3* mutations can cause postaxial polydactyly type A, preaxial polydactyly type IV and postaxial polydactyly type A/B. Until lately, it was supposed that there was no correlation between the site and the type of the mutation and the phenotype (Kalff-Suske et al. 1999), but recently it has been postulated that the site of frameshift and nonsense mutations can actually play a role in the determination of the syndrome. If the mutation happens in the first third of the *GLI3* gene it may cause GCPS, whereas mutations in the second third are associated with PHS. Some GCPS patients have been found to be mutated in the third part of the gene, however, no PHS patients are associated with changes in this section (Johnston et al. 2005).

Developmental defects, generally associated with limb deformities, can be caused also by mutations in other factors of the Shh pathway (figure 3-2). Thus, one might postulate that syndromes showing among other peculiarities changes in the shape or the number of digits should be caused by mutations in genes functionally related to the HH signaling cascade.

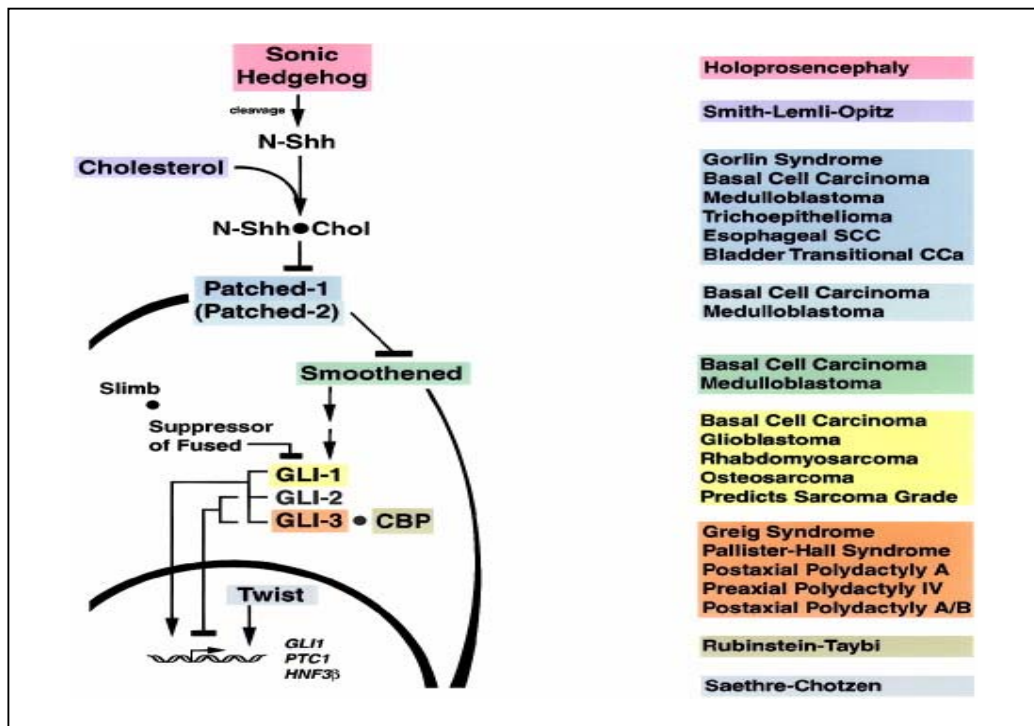


Figure 3-2: Shh pathway and associated syndromes

Diagram showing a shortened version of the vertebrate Shh pathway and a list of inherited diseases that are caused when the different parts of this pathway are mutated (Villavicencio et al. 2000).

3.4 Transcriptional regulation of genes of the Hh signaling cascade

Shh signaling governs the development of a long list of tissues and organs such as the limbs, the lungs, neural tube and brain, eyes, and muscles. It is present in a wide variety of normal adult tissues, including lung, colon, spleen, placenta, testis, myometrium and the formation of skeletal muscles (McDermott et al. 2005), (Warburton and Lee 1999), (Tyurina et al. 2005). However, little is known about the transcriptional regulation of genes of the Hh cascade. Most likely, a sophisticated system must exist that coordinates the up- and downregulation of the expression of different genes on different chromosomes in the correct spatiotemporal manner. For example, the expression of the members of the GLI family appears to be controlled by SHH (Marigo et al. 1996; Sasaki et al. 1999).

In the neural tube, Gli1 is restricted to the most ventral regions, Gli2 extends from the ventral region above the Gli1 domain to the most dorsal region, where the expression of Gli3 takes place. A similar expression pattern can be found in the limb buds

(Marigo et al. 1996; Buscher and Ruether 1998) where Gli1 is expressed posteriorly, Gli2 in the whole limb bud, and Gli3 anteriorly. Cis regulatory elements are expected to play a major role in the correct positional and temporal expression of the GLI family, orchestrating the different expression patterns and finally the polarizing activity of the SHH in these tissues.

As an example for such a regulatory element an enhancer for Shh, which is associated with preaxial polydactyly, has been identified 1 Mb upstream of the gene (Lettice et al. 2003). By mutating this region it has been shown that limb specific expression of Shh in mice is completely lost and degeneration of skeletal elements distal to the stylopod/zygopod junction is observed (Sagai et al. 2005).

3.5 Transcriptional regulation of coexpressed genes

Another question that arises concerning the transcriptional regulation is whether there are similar cis elements that regulate the expression of different genes in the same tissue. This problem was already addressed in 1986 when Buckingham and co-workers observed that the two murine myosin light chains (MLC1F and MLC3F), which are transcribed from two different promoters, as well as the α -cardiac actin gene have an enhancer element upstream of their promoters. This enhancer element and flanking nucleotide sequences is conserved between these genes. It was suggested that this whole structure is responsible for their correct spatiotemporal expression in the fetal skeletal muscle (Buckingham et al. 1986)

In 1995, Gomez-Skarmeta and colleagues identified a parallel regulation of two members of the AS-C complex, the proneural achaete (ac) and scute (sc) genes in *D. melanogaster*. Their simultaneous expression by groups of cells (the proneural clusters) located at constant positions in discs is orchestrated by enhancer-like elements distributed along a 90kb region of the AS-C. These elements could drive expression of the lacZ reporter gene only in specific proneural clusters (Gomez-Skarmeta et al. 1995).

3.6 Highly conserved regions outside of coding sequences

The situation encountered in SHH appears to be relevant for transcriptional control of many genes, in particular transcription factors. Spatiotemporal and quantitative regulation of eukaryotic gene expression can be governed by elements residing upstream, downstream or in the introns of the gene. These “long range cis regulatory elements”, that can be either enhancers or silencers, may exist far away for some genes that play key developmental roles. For example, an enhancer for *Shh* which is associated with preaxial polydactyly has been identified 1Mb upstream of the gene (Lettice et al. 2003). The significance of such long range elements is shown by the fact that they are conserved in many different species and they are associated with inherited diseases (Kleinjan and van Heyningen 2005).

3.7 Evolutionary sequence comparison facilitates the detection of long-range regulatory elements

Sequence comparison for identification of common transcriptional regulation, although not a recent idea, is a method being applied only recently to a greater extent. Novel computational programs and algorithms being developed can help to identify sequence similarity in transcription binding sites in order to better understand the regulation of coexpressed genes (Cora et al. 2005).

The complete sequence of man and other organisms, as well as the development of computational programs such as VISTA or PIPMAKER (Schwartz et al. 2000; Frazer et al. 2004) facilitates the genomic comparison of different species (figure 3-3). Conserved regions that reside outside of exons have been shown to contain functionally important elements such as cis regulators of transcription (Nobrega and Pennacchio 2004). By aligning genomes of evolutionary related species, such as human and other primates or mouse, one can find a large number of highly conserved regions. However, by comparing evolutionary distant species, such as human and the pufferfish *Fugu rubripes*, the conserved regions are narrowed down, potentially to the functionally most basic ones. Due to its very compact genome, which has lost much of most likely nonessential DNA and the presence of vertebrate specific orthologs for genes that have duplicated in the fish lineage, *fugu* is an excellent model for functional comparisons. It is ideal for identifying conserved elements of genes that

play important similar functional roles in many distantly related species. These elements can be easily tested for their function in transgenic systems such as zebrafish, *Xenopus*, and rodents (Venkatesh and Yap 2005).

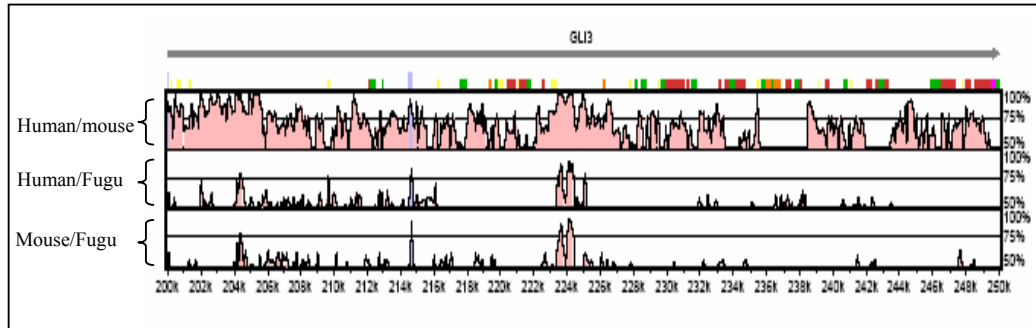


Figure 3-3: A vista plot depicting the conservancy of sequences.

The genomic sequence of the 5' part of the human *GLI3* gene (from exon 1 to intron 2) is compared with the corresponding sequence of mouse and fugu. The gray arrow shows the orientation of the gene, the blue boxes the 1st and the 2nd intron and the rest of the colored boxes stand for different repetitive elements. The higher the peak the higher the similarity of the sequences. By comparing sequences of evolutionary distant species, the conserved sequences are narrowed down to the ones that most probably contain elements essential for the function of the gene throughout different organisms. Regions highly conserved even in fugu (>75%) appear in the 1st and 2nd exon of the *GLI3* gene as well as in introns.

3.8 Transcriptional regulation of paralog genes

Paralog genes originate from duplication steps during the evolution of genomes. Such events result from duplication events within chromosomes (HOX cluster, hemoglobin and keratin clusters) from translocation of duplicated elements to other chromosomes, and/or from polyploidization phenomena that repeatedly occurred during evolution. They generated gene families, members of which are structurally and frequently still functionally related. Depending on the time point when these events occurred, similarities might apply to nongenic sequence, as well. Coding sequences, which were inherited as separate entities following a duplication step early in evolution, should only retain high sequence similarity if there are strong functional restraints. Whereas these restraints may have conserved protein sequence similarity between orthologs from pro- to eukaryotes, paralog genes might have developed sequence and functional divergency over time.

The classical model for the creation of paralogs suggests that originally the duplicated genes have identical roles. Mutations can either turn one of the two genes into a pseudogene, if they are deleterious, or rarely give novel attributes to it, if they are beneficial, while the other copy preserves the initial function (Ohno 1970).

Duplication-degeneration-complementation (DDC) process is another model trying to explain the creation of paralog genes (figure 3-4). By this, one of the two genes can become nonfunctional by accumulation of null mutations in the coding or regulatory region, or it can get a new function by mutations, which will be preserved by the natural selection. Finally, degenerative mutations can occur for different subfunctions in the two genes, thus the combined action of both of them is needed for the normal function of the cell. In this way, the paralog genes are preserved during evolution (Force et al. 1999).

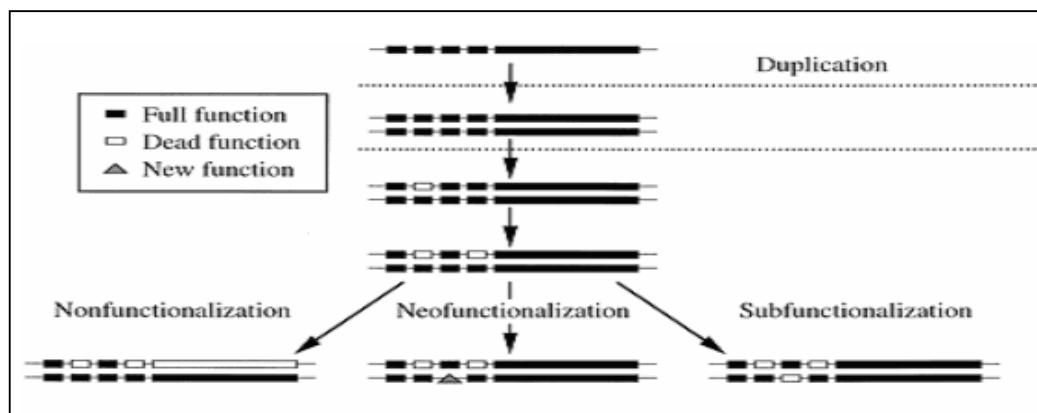


Figure 3-4: The DDC model for the development of paralog genes (Force et al. 1999).

Duplicated genes have three different possible fates. One copy can completely lose its activity due to mutations in the coding regions or in its regulatory elements, and thus become a pseudogene. Mutations on regulatory regions can also give one ortholog a new function, and if this change happens at the expense of an otherwise essential function, both copies are preserved. Finally, different mutations on the two copies can lead to the necessity of the combined function of both copies so that the function of the original gene is preserved. The long boxes represent the coding region whereas the small ones depict the regulatory sequences.

If, however, as is the case for GLI genes, the paralogs continue to maintain a coordinated expression pattern, one might expect that in addition to similarities in the

coding parts of the genes, regulatory, non-coding DNA sequence elements might be conserved, as well.

Promoter sequences of such genes appear to be different, and there have not been many reports of common homologous, potentially cis regulatory regions between paralogs, in addition to the coding sequences. One example of transcriptional regulation by conserved enhancers can be found for *Hoxa3* and *Hoxb3*. Both genes contain in their enhancers binding sites for Krml-1 but in different copy numbers, resulting in similarities and differences in the expression pattern of *Hoxa4* and *Hoxb3* in mice hindbrain (Manzanares et al. 1999). *Hoxa4*, *Hoxb4* and *Hoxd4* also share enhancer elements for mesodermal and neural regulation. These regulators are conserved also in the orthologs between human and mouse (Morrison et al. 1997). The NDP kinase family also shows conservation of regulatory regions but with varying binding motifs, explaining the spatiotemporal differences in the expression of the different paralogs of this family (Ishikawa et al. 2003). Five cis-acting elements (SRE, MyoD binding site, Elk-1 binding site, positive and negative YY1 binding sites) have been found to regulate six actin paralog genes, the beta and gamma cytoplasmic actins, alpha and gamma smooth muscle actins and alpha skeletal and alpha cardiac actins. Although one or more of these elements are present in all of the paralogs, their numbers, sequences and distribution patterns vary remarkably. Similar distribution patterns of elements can be found in coexpressed paralogs. This means that the sequence and the architecture of the cis regulatory elements can dictate the tissue and temporal specificity of the expression of different gene family members. (Liu et al. 2000).

These examples show that paralog genes can share common regulatory elements for overlapping expression. Still, differences in their placement and usage might allow distinct functions as well.

3.9 Objectives of the study

The aim of this thesis was to contribute to the detection and functional characterization of cis-acting regulatory elements for *GLI3* and their potential relevance for pathogenity in man by answering the following questions:

- How is the expression of human *GLI3* regulated?

- Do *GLI2* and *GLI3* share similar regulatory elements?
- Are mutations in *GLI3* regulatory elements involved in pathogenicity?

Towards this end, studies were undertaken to enlarge the list of cases with *GLI3* morphopathies, which cannot be attributed to mutation in coding sequences of the gene. These cases are candidates to be searched for mutations in cis-regulatory elements upstream, downstream or in the introns of the gene.

The transcriptional control of *GLI3* gene expression involves promoter as well as cis-acting enhancer sequences. As a first step, a minimal promoter was to be defined and tested functionally as well as by mutagenesis. Factors potentially binding to this minimal promoter and specifically governing *GLI3* expression were to be identified.

Functional studies in transgenic mice suggest that *GLI2* and *GLI3* might have greatly overlapping functions, however, sequence comparison of human and murine *GLI2* suggested, that human *GLI2* might not act in a similar way. To prepare for a functional comparison of human *GLI2* and *GLI3* expression, it was a task of this study to search the DNA and putative coding sequences for homologies.

The focus in the search for conserved cis-regulatory elements was set on *GLI3*. Evolutionary genomic sequence comparisons were to be applied to guide the search for potential enhancers. Predicted regions had to be tested for their regulatory capacity on reporter genes with foreign as well as the original *GLI3* minimal promoter in transiently transfected cell cultures as well as in transgenic mice. In mouse embryos it was to be determined, if the expression patterns follow the time course and the spatial pattern of the endogenous mouse *GLI3*. Mutagenesis of potential enhancers followed by tests for retention of their regulatory capacity should pinpoint particularly critical sequences, and potentially binding transcription factors specifically regulating *GLI3* expression should be identified.

The answers should contribute to our understanding of the spatial and temporal control of the expression of a key factor of the hedgehog signaling cascade and provide insight into the potential role of highly conserved non-coding sequence elements in the human genome.

4 MATERIALS AND METHODS

4.1 MATERIALS

4.1.1 Devices and accessories

12 Well Plate Cellstar	Greiner bio-one, Frickenhausen
ABI Prism 310 Genetic Analyzer	Applied Biosystems, USA
ABI Prism 377 DNA-Sequencer	Applied Biosystems, USA
Autoclave	Grössner, Hamburg
AutoLumat LB 953	Berthold, Bad Wildbad
Balance AE 240	Mettler, Switzerland
Balance PM 2000	Mettler, Switzerland
Cell Counter	Neubauer, Marienfeld
Celloshaker Variospeed	Biotec Fischer, Reiskirchen
Centrifuge 5415C	Eppendorf, Hamburg
Centrifuge 5415 R	Eppendorf, Hamburg
Centrifuge 5417C	Eppendorf, Hamburg
Centrifuge Labofuge 400	Heraeus, Hanau
Centrifuge Sorvall GLC-2B	Du Pont, Dreieich
Centrifuge Sorvall RC-5b	Du Pont, Dreieich
Centrifuge Sorvall RT6000	Du Pont, Dreieich
Eissmaschine ZBE 70-35	Ziegra, Isernhager
Electrophoresis Hoefer apparatus SE 600	Hoefer Pharmacia Biotech, USA
Electrophoresis Horizontal tank A2	Owl Scientific Inc. Wobum, USA
Electrophoresis Horizontal tank B14	Owl Scientific Inc. Wobum, USA
Falcon Tubes 14ml	Falcon, USA
Falcon Tubes 15ml	Falcon, USA
Falcon Tubes 50ml	Falcon, USA
Gel Documentation System E.A.S.Y. RH-3	Herolab, ST. Leon Rot, Wiesloch
Incubator Cytoperm	Heraeus, Hanau
Incubator shaker	New Brunswick Scientific, USA
Incubator type B5042 E	Heraeus, Hanau
Laminair HA 2448 GS	Heraeus, Hanau

Laminair Microflow	Stolco, Düsseldorf
Micro-Shaker	Flow Laboratories, Meckenheim
Microwave Oven	Bosch, Gerlingen-Schillerhöhe
Minisart filter NML 0.45µM	Sartorius GmbH, Göttingen
Multitemp II Thermostatic Circulator 2219	LKB Bromma
Nunc Cryotube Vials	Nalge Nunc International, Neerijse, Belgium
Optic Microscope Diavert	Leitz, Wetzlar
PCR Strip Tubes	Starlab, Ahrensburg
PH Indicator Paper Neutralit	Merck, Darmstadt
Photometer GeneQuant II	Pharmacia Biotech, Uppsala, Sweden
Pipette Tips Biosphere 10µl	Sarstedt, Nümbrecht
Pipette Tips Biosphere 200µl	Sarstedt, Nümbrecht
Pipette Tips TipOne 101-1250µl	Starlab, Ahrensburg
Pipettes:	
Eppendorf Pipettes	Eppendorf, Hamburg
Electric pipette Accu Jet	Brand, Wertheim
Hamilton Pipette	Hamilton, Bonaduz, Sweden
Power Supply EPS 500/400	Pharmacia, Uppsala, Sweden
Power Supply LKB ECPS 3000/150	Gibco, BRL, USA
Power Supply Powerpad 200	BioRad, USA
Reaction tubes 3810 1,5ml	Eppendorf, Hamburg
Safe-lock tubes 0,5ml	Eppendorf, Hamburg
Sarstedt tube 5ml	Sarstedt, Nümbrecht
Shaker RT-1S, Rocky	Kobe, Marburg
Syringe needle 20 gauge	BD, Heidelberg
Tissue Culture Flasks 250ml Cellstar	Greiner bio-one, Frickenhausen
Thermocycler I cycler	BioRad, USA
Thermocycler Perkin Elmer 2400	Applied Biosystems, USA
Thermocycler Perkin Elmer 9600	Applied Biosystems, USA
Thermocycler Trio Thermoblock	Biometra, Göttingen
Vortex VF2	IKA Labortechnik, Staufen
Waterbath type 1086	GFL, Grossburgwedel

Waterbath type 1083	GFL, Grossburgwedel
Whatman 3M filter paper	Whatman Lab Sales, USA

4.1.2 Chemicals

All chemicals that are not described in this list were purchased from Gibco BRL (USA), Merck (Darmstadt), Riedel-de-Haen (Seelze), Roche Diagnostics (Mannheim), Roth (Karlsruhe), Serva (Heidelberg), Sigma (München)

40% Acrylamid	Pharmacia Biotech, Freiburg
2% Methylenbisacrylamid	Pharmacia Biotech, Freiburg
Agar	Gibco BRL, USA
Agarose	Invitrogen, Groningen, Netherlands
Bacto Yeast Extract	BD, Le Pont de Claix, France
Bactotryptone	BD, Le Pont de Claix, France
Bromophenol Blue	Merck, Darmstadt
Coomassie Blue	Carl Roth, Karlsruhe
Cryoprotective medium	Cambrex, Verviers, Belgium
DMSO (Dimethylsulfoxide)	Serva Reinbiochemica, Heidelberg
D-PBS	Gibco BRL, USA
EDTA	Diagnostic International, USA
Fetal Calf Serum	Boehringer Mannheim, Mannheim
Formamid	Merck, Darmstadt
Glycine	Riedel de Haen, Seelze
L-Glutamine	Cambrex, Verviers, Belgium
Long Ranger Gel Solution	BioWhittaker Molecular Applications, USA
MEM Non Essential Aminoacids	Gibco BRL, USA
Penicillin/Streptomycin	Gibco BRL, USA
Phenol Red	Serva, Heidelberg
Rotiphorese ® NF-Acrylamide/Bis	Carl Roth, Karlsruhe
RPMI 1640 Medium – L Glutamine	Gibco BRL, USA
SDS	Merck, Darmstadt
Temed	Carl Roth, Karlsruhe

Tris – Cl	Carl Roth, Karlsruhe
Triton X-100	Serva, Heidelberg
Tween-20	Sigma, Deisenhofen
KOH	Merck, Darmstadt
MOPS	Sigma, Deisenhofen
NaCl	Riedel de Haen, Seelze
NaOH	Merck, Darmstadt
Water -for PCR, Sequencing: HPLC	Merck, Darmstadt
-for SSCA: Ampuwa	Fresenius, Oberursel

4.1.3 Buffers and standard solutions

All buffers and solutions were made with MilliQ water.

4.1.3.1 DNA Isolation

P1	15 mM Tris/HCl (pH 8.0) 10 mM EDTA 100 µg/ ml RNase A
P2	200 mM NaOH 1% SDS (w/v)
P3	3 M Potassium acetate, pH 5.5
QBT	750 mM Sodium chloride 50 mM MOPS (pH 7.0) 15% isopropanol (v/v) 0.15% Triton X-100 (v/v)
QC	1 M Sodium chloride 50 mM MOPS (pH 7.0) 15% isopropanol (v/v)
QF	1.25 M Sodium chloride 50 mM Tris/HCl (pH 8.5)

	15% Isopropanol (v/v)
TE	10 mM Tris-Cl (pH 8.0) 1 mM EDTA
TES	10 mM Tris-Cl, pH 7,5 1 mM EDTA, pH 7,5 0.1% (w/v) SDS

4.1.3.2 Gel Electrophoresis

50x TAE	2 M Tris base 5.71% Acetic acid (v/v) 50 mM EDTA (pH 8)
---------	---

4.1.3.3 Bacterial culture

All media are autoclaved for 20min before use.

LB Medium	1% Bactotryptone (w/v) 0.5% Yeast Extract (w/v) 1% NaCl (w/v)
LB Agar	1% Bactotryptone (w/v) 0.5% Yeast Extract (w/v) 1% NaCl (w/v) 1.5% Agar (w/v)
DYT-Glycerin	1.6% Bactotryptone (w/v) 1% Yeast Extract (w/v) 0.5% NaCl (w/v) 80% Glycerin (v/v)

4.1.3.4 Cell culture

RPMI Medium 4% or 10%	90% RPMI Medium (v/v) 1% L-Glutamine (v/v)
-----------------------	---

4% or 10% Fetal Calf Serum (v/v)
1% Non Essential Aminoacids (v/v)
2% Penicillin/Streptomycin (v/v)

PBS + EDTA pH 7,5

0.8% NaCl (w/v) (Merck)
0.02% KCl (w/v) (Merck)
0.14% Na₂HPO₄ (w/v) (Merck)
0.02% KH₂PO₄ (w/v) (Merck)
0.02% EDTA (w/v) (Roth)
0.1% Phenol red (v/v) (Serva)

4.1.4 Enzymes

BD Powerscript Reverse Transcriptase

BD Biosciences,
Heidelberg

Taq DNA polymerase

Promega, Madison, WI,
USA

Herculase enhanced DNA polymerase

Stratagene, Cedar Creek,
USA

PfuTurbo DNA polymerase

Stratagene, Cedar Creek,
USA

Klenow polymerase

MBI Fermentas, St Leon-
Rot

Calf Intestinal Alkaline phosphatase

NEB, Beverly, USA

Rnase A

Qiagen, Hilden

T4 DNA ligase

Invitrogen, Groningen,
Netherlands

4.1.5 Loading dye

6x Agarose Gel Loading Dye

2.5 mg/ml Bromophenol
blue

4.1.6 Primers

4.1.6.1 Primers for SSCA

Exon	Size of exon in bp	Primer	5'-3' primer sequence	Formamide	Annealing temperature	Length of PCR product	Position in cDNA or intron sequence
II	166	ExIIfor1	AGA TGA CAG GTG AAG TGG AC	-	60°C	211	5' intron
		ExIIrev1	CTC GCT CAC ATC TGT TCG AGT G				Until nt 39
		ExIIfor2	GAA GAC ATC ATG GAG GCC CAG	-	66°C	217	From nt -10
		ExIIrev2	CGC TCA ATT CAC AAG GAA TGC				3' intron
III	243	ExIIIfor1	CGA GAA TGA GAC CTA ATT GAT	-	55°C	293	5' intron
		ExIIIrev1	CAT GGC AAA CAC CGT CCC				Until nt 333
		ExIIIfor2	CTT CAA CAT CGA GTG ACG AG	+	54°C	185	From nt 224
		ExIIIrev2	CCA TAC CTC CTG AAC AAG TG				3' intron
IV	106	ExIVfor	GTT GCT TTG TGA ATC GGA ATG	-	58°C	226	5' intron
		ExIVrev	AGC CAG CAT CTC GTT CCA TTT C				3' intron
V	206	ExVfor1	GAG ACA GCC ATA GCC TGT G	-	61°C	215	5' intron
		ExVrev1	GAG CGG ATA TAG TCC ATG TAG				Until nt 617
		ExVfor2	CTA CGT ATC CGG ACC TGC	-	63°C	228	From nt 497
		ExVrev2	AGG AGG CTG CAT GAT CTC				3' intron
VI	147	ExVIfor	GAC TCT GAG ATG CCT CAA GA	-	62°C	264	5' intron
		ExVIrev*	GCC ATT TCC CAA GAC TC				3' intron
VII	202	ExVIIfor1	TGG TAT AGA GTC CAA CTG CA	-	64°C	215	5' intron
		ExVIIrev1*	GCT GAA GAG CTG CTA CGG				Until nt 989
		ExVIIfor2	GTC CAT ATC ACC ACT CTC CG	-	56°C	205	From nt 901
		ExVIIrev2	GCA AAC AAG TGC TGA CAT TA				3' intron
VIII	214	ExVIIIfor1	GTG TAC TTG TCT GAG CTC AGC	-	62°C	248	5' intron
		ExVIIIrev1	CTG GAA TAG GCC TCT GTG TTG G				Until nt 1183
		ExVIIIfor2	CTC TCT CCA CAT GCA TCA GCA	-	66°C	216	From nt 1065
		ExVIIIrev2	CAA CTG ACG TGG TGG CCT GC				3' intron
IX	114	ExIXfor	TGC CCG AGT TGG CAA GG	-	61°C	265	5' intron
		ExIXrev	GAC CCA AAG ACA CCA GTC TTG				3' intron
X	141	ExXfor	CCT CCT GTT GTG TCT GAT TC	-	55°C	247	5' intron
		ExXrev	GTC ACA GAG CTG TAA AGC TCG				3' intron
XI	150	ExXIfor*	TGA TGA ATA CGT TTC CAT TTG	+	51°C	231	5' intron
		ExXIrev*	AAG GAC CCA AGT GTG CCT G				3' intron
XII	164	ExXIIfor	CGT GTT AGG CAT AAG CTG CTG	-	59°C	265	5' intron
		ExXIIrev*	CCT TAT GCA AGC TCC ATG CC				3' intron
XIII	291	ExXIIIfor1	CTT TTG GGA TGG GGC ATA GTC	-	64°C	246	5' intron
		ExXIIIrev1*	GAC CTG GAC TGT GAA TGG CTG				Until nt 1991
		ExXIIIfor2	GCC CAG AGG CTC ATG TCA CC	-	65°C	243	From nt 1898
		ExXIIIrev2	GGC CTG CTG TGA AGT CAG AAG				3' intron

XIV	328	ExXIVfor1	CAT GAG CTC TCA TTA TTG GCT	-	60°C	239	5' intron
		ExXIVrev1	GTG GAA ATG GTT GAG TCC ATG				Until nt 2267
		ExXIVfor2	CAT CAG CAA CTA TTC CAA CAG	-	60°C	223	From nt 2157
		ExXIVrev2	CTA TGA GAG GAG AGA CCG CAG				Until nt 2425
		ExXIVfor3	CGG CAG GGA CCA AAT GGA TG	-	60°C	223	From nt 2306
		ExXIVrev3	GGA CTG GTG GAG AAA CTA				3' intron
XV	2570	ExXVfor1	GTG TAA GTC ATG CTC GAA GGG	+	61°C	198	5' intron
		ExXVrev1	GAG CTG ATG GTG CTG GCG C				Until nt 2567
		ExXVfor2	CAA CAC CTG CAG CTT GGG TG	+	57°C	224	From nt 2445
		ExXVrev2	CTC ACG TTC TGC GGC CGG C				Until nt 2669
		ExXVfor3	GCA GCC GCC GCT CCT CAG G	+	62°C	214	From nt 2579
		ExXVrev3	GCC TTG AGG CGG TAC TGC TG				Until nt 2792
		ExXVfor4	CAT CTC CAC CGA CGC CTC	+	60°C	213	From nt 2691
		ExXVrev4	CAC GCC AGG CTC GAG GGC				Until nt 2904
		ExXVfor5	CAA GTA CGC GGC TGC CAC	+	57°C	228	From nt 2793
		ExXVrev5	GTC GCT GGC CCT CCT CAC G				Until nt 3021
		ExXVfor6	CCG AGG AGG TGC AGC GAC G	+	65°C	228	From nt 2926
		ExXVrev6	GCC GCG TGT AAT TCT GAA GCA C				Until nt 3154
		ExXVfor7	CTC GTG TGC CGC GCT TCA G	+	62°C	238	From nt 3056
		ExXVrev7	CTG CAC CAC GTC GTC CGG C				Until nt 3294
		ExXVfor8	CAG CAT CAC CGA GAA CGT CAC	+	63°C	262	From nt 3201
		ExXVrev8	TGC TGC CCT CGG GGC AGG				Until nt 3463
		ExXVfor9	CCA CGG GCC CGG TGA CTT TG	+	64°C	262	From nt 3369
		ExXVrev9	CAG CAG GCC CGC TCC TCA AG				Until nt 3631
		ExXVfor10	CAA GTG TGG GCC GCG GCC	-	64°C	254	From nt 3531
		ExXVrev10	GGC TGC CTG TTG AGA CAG TTC				Until nt 3785
		ExXVfor11	CTT GAT GCT CCA CAA CAG CC	+	63°C	243	From nt 3684
		ExXVrev11	CAC CAT GCT GCC AGC TGA CTC				Until nt 3927
		ExXVfor12	GCC TCA AAG CTG AAG AGC AC	+	57°C	244	From nt 3832
		ExXVrev12*	CTT GGT AGA TGT TGA TGT GTG				Until nt 4075
		ExXVfor13	CGA CAG CAT GCA GCA CCC	+	62°C	271	From nt 3997
		ExXVrev13	CAT TCA CCC TGC AGG TCT G				Until nt 4258
		ExXVfor14	CAG CCA TGT GCC AGC TTT G	-	68°C	260	From nt 4150
		ExXVrev14	CTG TAG CAG GCA GCT GGC G				Until nt 4410
		ExXVfor15*	CTA TGA CCA AAC CGT GGG C	-	61°C	261	From nt 4329
		ExXVrev15	GTT CTG AAT GAT ACT TGG GC				Until nt 4590
		ExXVfor16	GGT ACA GAT TGA CTT CGA TGC	-	61°C	276	From nt 4503
		ExXVrev16*	GAT TTC CGT TGG TTG CAG TC				Until nt 4779
		ExXVfor17	GAG TTC TTT GCT GAC CTC	+	50°C	247	From nt 4687
		ExXVrev17	CTC CCT AGA ATA CTT TAG G				3' UTR

4.1.6.2 Primers for cloning

Primer	Annealing temperature	5'-3' primer sequence	Restriction site tag
Ex1cut for	80.4 °C	GCCTCGAGTGGCGGTGTGGAGAGAGAAGAG	XhoI
Ex1cut rev	80.1 °C	GCCTCGAGGCTTTCTCGCTCCTTCCAGG	XhoI
Ex1cut forwnew	81.2 °C	GCCTCGAGCAGACGAGTCCGGTGGAGAG	XhoI
Ex1prom Atrunc F	80.6 °C	GCCTCGAGGCAGGCATGAATGTGTCTGTG	XhoI
Hom2New F	77.6 °C	GCAAGCTTGAGCAATTGCAGAGTGCAGG	HindIII
Hom2New R	74.3 °C	GCAAGCTTCACCTCTCTCCAACCAGCTAG	HindIII
Hom3forwnew	79.1 °C	GCAAGCTTGAATGTCGCAGGGCAGAAAATG	HindIII
Hom3 rev	72.9 °C	GCAAGCTTCATGGTGCCTTGTACAGGTG	HindIII
3'cutHom3rev	78.1 °C	GCGGTACCTGAACAACAGCAGCAGACC	KpnI
5'cutHom3for	75 °C	GCGGTACCCAAGACAATCACAGTAGTGGTG	KpnI
Hom4 forwnew	73.3 °C	GCAAGCTTCATCAACGATATGGTGCAG	HindIII
Hom4 rev	69.9 °C	GCAAGCTTGTCTGTAATTGCAGTTGTTC	HindIII
Hom3KpnI for	81.4 °C	GCGGTACCGAATGTCGCAGGGCAGAAAATG	KpnI
Hom3KpnI rev	78.2 °C	GCGGTACCCATGGTGCCTTGTACAGGTG	KpnI
Hom4KpnI for	76 °C	GCGGTACCCATCAACGATATGGTGCAG	KpnI
Hom4KpnI rev	72.8 °C	GCGGTACCGTCTGTAATTGCAGTTGTTC	KpnI
Hom2NewKpnI F	80.1 °C	GCGGTACCGAGCAATTGCAGAGTGCAGG	KpnI
Hom2NewKpnI R	76.8 °C	GCGGTACCCACCTCTCTCCAACCAGCTAG	KpnI
Ex1cutp1230for	81.2 °C	GCCTGCAGTGGCGGTGTGGAGAGAGAAGAG	PstI
Ex1cutp1230rev	80.9 °C	GCCTGCAGGCTTTCTCGCTCCTTCCAGG	PstI

4.1.6.3 Primers for mutagenesis

Primer	Annealing temperature	5'-3' primer sequence	Restriction site
Hom2mut1 F	78.6 °C	GACACTGCGGGAGGATCCAGGACTTTGAAAC	BamHI
Hom2mut1 R	78.6 °C	GTTTCAAAGTCCTGGATCCTCCCGCAGTGTC	BamHI
Hom2mut2 F	76.9 °C	GACTTTGAAACTTGAGGATCCAGAGCTTGCTT TCAAC	BamHI
Hom2mut2 R	76.9 °C	GTTGAAAGCAAGCTCTGGATCCTCAAGTTTCA AAGTC	BamHI
Hom2mut3 F	78.4 °C	CCTCAAAAGCTAGGGATCCAGGGCTCTGAAA TTTG	BamHI
Hom2mut3 R	78.4 °C	CAAATTTAGAGCCCTGGATCCCTAGCTTTTG AGG	BamHI
Prommut1 F	88.0 °C	GAGGGCGCGGGTAAGCTTACGCCC GG GCCC	HindIII
Prommut1 R	88.0 °C	GGGCCCCGGGCGTAAGCTTACCCGCGCCCTC	HindIII
Prommut2 New F	87.8 °C	CCTCGCCGGGTGAATGGGGCTTCGGCG	-
Prommut2 New R	87.8 °C	CGCCGAAGCCCCATTACCCGGCGAGG	-
Prommut3 New F	93.2 °C	GGCGCGCCGGGACCGCGCGGGTCAAG	-
Prommut3 New R	93.2 °C	CTTGACCCCGCGCGGTCCCGGCGCGCC	-
Prommut4 F	81.4 °C	CTCCTCCTCCTCCGCAAGCTTGTCTTCACACA GAC	HindIII
Prommut4 R	81.4 °C	GTCTGTGTGAAGACAAGCTTGC GGAGGAGGA GGAG	HindIII
Prommut 5New F	79.2 °C	GTCAAGTAGGGGACGAGGAGGAGCAGGC	-
Prommut 5New R	79.2 °C	GCCTGCTCCTCCTCGTCCCCCTACTTGAC	-
InrmutF	79.6 °C	CAGCTGGGACCTTCTCCGGGCTTGG	-
InrmutR	79.6 °C	CCAAGCCCGGAGAAGGTCCAGCTG	-

4.1.6.4 Primers for RACE PCR

Primer	Annealing temperature	5'-3' primer sequence
AP1	71 °C	CCATCCTAATACGACTCACTATAGGGC
AP2	77 °C	ACTCACTATAGGGCTCGAGCGGC
GLI2Race2	75.2 °C	GCTTCTTGCTTCTCGGAGGCAGTGG
GLI2Race3	81 °C	GCTCATCGTCCTGGGTGGCAATCCTTGC
GLI2Race4	79 °C	GCAGACGTCTCCATCTCAGCCGCTCATC
GLI3RT1	77.7 °C	CATTAGAAGTGGTGCTGGAGGCAACGGC
GLI3RT2	76.7 °C	CTCGCTCACATCTGTTCGAGTGGAGCAC

4.1.6.5 Primers for primer extension

Restriction site tag	Annealing temperature	5'-3' primer sequence	Modification
Exten	63.3 °C	TTTGTCTTCTCGCTCCTTC	5' IRD700
Lucexen	62.7 °C	TCCATCTTCCAGCGGATAG	5' IRD700

4.1.6.6 Primers for sequencing

Primer	Annealing temperature	5'-3' primer sequence
GLI2spec F	71 °C	CACCAAGAAGCAGCGCAATGAC
GLI2spec R	68.8 °C	TTCTCCTTCTTGAGCTGCTCGAAC
Rvprimer3	56.5 °C	CTAGCAAAATAGGCTGTCC
GLprimer2	63.3 °C	CTTTATGTTTTTGCGTCTTCC
M13 F	57.5 °C	GTAAAACGACGCCAG
M13 R	50.6 °C	CAGGAAACAGCTATGAC
NFATP F	81.8 °C	GCGAATTCCTGCTGGTCCGCCCACTTG
NFATP R	82 °C	GCGAATTCCTGTTGCTGGTAGCGGGCGTC
b-actin F	70.2 °C	GACTTCGAGCAAGAGATGGCCAC
b-actin R	61.4 °C	CTTCATGATGGAGTTGAAGGTAG
LacZ F	71.8 °C	CAACAGTTGCGCAGCCTGAATG
LacZ R	72.3 °C	GTGGGAACAAACGGCGGATTG

4.1.7 DNA size standards

100 bp DNA ladder

λ/HindIII/ EcoRI

Gibco BRL, Eggenstein

MBI Fermentas, St Leon-

Rot

4.1.8 Reaction kits

BigDye Terminator v1.1 Cycle Sequencing Kit	Applied Biosystems, Warrington, UK
Dual-Luciferase Reporter Assay System	Promega, Madison, WI, USA
DYEnamic ET Terminator Cycle Sequencing Kit	Amersham, Buckinghamshire, UK
Effectene Transfection Reagent	QIAGEN, Hilden
Human Total RNA Master Panel II	BD, Le Pont de Claix, France
Marathon-Ready cDNA	BD, Le Pont de Claix, France
QIAfilter Plasmid Maxi Kit	Qiagen, Hilden
QIAGEN Plasmid Midi Kit	Qiagen, Hilden
QIAprep Spin Miniprep Kit	Qiagen, Hilden
QIAquick Gel Extraction Kit	Qiagen Hilden
QIAquick PCR Purification Kit	Qiagen, Hilden
QuikChange Site Directed	
Mutagenesis Kit	Stratagene, Cedar Creek, US
Ready-To-Go- PCR Beads	Amersham Pharmacia Biotech, Piscataway, USA
RNeasy MinElute Cleanup Kit	Qiagen, Hilden
TOPO TA Cloning Kit	Invitrogen, Groningen, Netherlands

4.1.9 Plasmids used in this work

Nr.	Name	Comment	Reference
1	pGLI2 N ²² 22	5' upstream region of already published cDNA human GLI2 sequence, TA cloned in PCR 2.1 vector, 1379 bp size of insert	This work
2	pGL3basic-1.9 kb StuI-DpnI-luc	1.97 kb region containing 1,4 kb upstream of Exon1, Exon1 and 0.5 kb downstream of Exon1 of human GLI3, cloned in Sma-I site of pGL3basic vector in (+) orientation	This work
3	pGL3basic-3.7 kb StuI-StuI-luc	3.7 kb region containing 1,4kb upstream of Exon1, Exon1 and 2,2 kb downstream of Exon1 of human GLI3, cloned in Sma-I site of pGL3basic vector in (+) orientation	This work
4	Hom3a(-) – p1230	Hom3 region (386 bp) cloned in HindIII site of p1230 vector in (-) orientation. A is exchanged to T in 1 st nt of Hom3 region.	This work
5	p1230	The b-globin promoter/lacZ reporter gene- fragment was inserted into the SmaI site of pBluescript II KS vector. B	This work

		globin promoter is residing downstream of the T3 promoter.	
6	Hom3a(+) – p1230	Hom3 region (386 bp) cloned in HindIII site of p1230 vector in (+) orientation. No mutation.	This work
7	Hom4a(+) – p1230	Hom4 region (172 bp) cloned in HindIII site of p1230 vector in (+) orientation. A is deleted in 6 th nt of Hom4 region.	This work
8	Hom4a(-) – p1230	Hom4 region (172 bp) cloned in HindIII site of p1230 vector in (-) orientation. No mutation.	This work
9	pGL3basic-promGLI3-600-luc	686 bp region containing 619 bp upstream of presumed initiation of human GLI3 as well as exon1, cloned in (+) orientation in XhoI site of pGL3basic vector	This work
10	Hom3a(-) – p1230	Hom3 region (386 bp) cloned in HindIII site of p1230 vector in (-) orientation. No mutation.	This work
11	Hom4a(+) – p1230	Hom4 region (172 bp) cloned in HindIII site of p1230 vector in (+) orientation. No mutation.	This work
12	pGL3basic-promGLI3-300-luc	425 bp region containing 358 bp upstream of presumed initiation of human GLI3 as well as exon1, cloned in (+) orientation in XhoI site of pGL3basic vector.	This work
13	Hom2New (-)	Hom2 region with more information (434 bp) cloned in HindIII site of p1230 vector in (-) orientation. No mutation.	This work
14	pGL3basic-Hom3(+)-luc	Hom3 region (386 bp) cloned in HindIII site of pGL3basic expression vector in (+) orientation. No mutation.	This work
15	pGL3basic-Hom3(-)-luc	Hom3 region (386 bp) cloned in HindIII site of pGL3basic expression vector in (-) orientation. No mutation.	This work
16	pGL3basic-Hom3(+)-promGLI3-300-luc	Hom3 region (386 bp) cloned in KpnI site of pGL3basic-promGLI3-300-luc construct in (+) orientation. No mutation.	This work
17	pGL3basic-Hom3(-)-promGLI3-300-luc	Hom3 region (386 bp) cloned in KpnI site of pGL3basic-promGLI3-300-luc construct in (-) orientation. No mutation.	This work
18	pGL3basic-Hom4(+)-luc	Hom4 region (172 bp) cloned in HindIII site of pGL3basic expression vector in (+) orientation. No mutation.	This work
19	pGL3basic-Hom4(-)-luc	Hom4 region (172 bp) cloned in HindIII site of pGL3basic expression vector in (-) orientation. No mutation.	This work
20	pGL3basic-Hom4(+)-promGLI3-300-luc	Hom4 region (172 bp) cloned in KpnI site of pGL3basic-promGLI3-300-luc construct in (+) orientation. No mutation.	This work

21	pGL3basic-Hom4(-)-promGLI3-300-luc	Hom4 region (172 bp) cloned in KpnI site of pGL3basic-promGLI3-300-luc construct in (-+) orientation. No mutation.	This work
22	pGL3basic-Hom2(+)-luc	Hom2 region with more information (434 bp) cloned in HindIII site of pGL3basic vector in (+) orientation. No mutation.	This work
23	pGL3basic-Hom2(-)-luc	Hom2 region with more information (434 bp) cloned in HindIII site of pGL3basic vector in (-) orientation. No mutation.	This work
24	pGL3basic-Hom2(+)-promGLI3-300-luc	Hom2 region with more information (434 bp) cloned in KpnI site of pGL3basic-promGLI3-300-luc construct in (-) orientation. No mutation.	This work
25	pGL3basic-Hom2(-)-promGLI3-300-luc	Hom2 region with more information (434 bp) cloned in KpnI site of pGL3basic-promGLI3-300-luc construct in (-) orientation. No mutation.	This work
26	Hom2New (+)	Hom2 region with more information (434 bp) cloned in HindIII site of p1230 vector in (+) orientation. No mutation.	This work
27	pGL3basic-3'Hom(+)-luc	Hom3 region lacking part of its 5' sequence (273 bp) cloned in KpnI site of pGL3basic vector in (+) orientation. No mutation.	This work
28	pGL3basic-3'Hom(-)-luc	Hom3 region lacking part of its 5' sequence (273 bp) cloned in KpnI site of pGL3basic vector in (-) orientation. No mutation.	This work
29	pGL3basic-3'Hom3(+)-promGLI3-300-luc	Hom3 region lacking part of its 5' sequence (273 bp) cloned in KpnI site of pGL3basic-promGLI3-300-luc construct in (+) orientation. No mutation.	This work
30	pGL3basic-3'Hom3(-)-promGLI3-300-luc	Hom3 region lacking part of its 5' sequence (273 bp) cloned in KpnI site of pGL3basic-promGLI3-300-luc construct in (-) orientation. No mutation.	This work
31	pGL3basic-5'Hom3(+)-promGLI3-300-luc	Hom3 region lacking part of its 3' sequence (267 bp) cloned in KpnI site of pGL3basic-promGLI3-300-luc construct in (+) orientation. No mutation.	This work
32	pGL3basic-5'Hom3(-)-promGLI3-300-luc	Hom3 region lacking part of its 3' sequence (267 bp) cloned in KpnI site of pGL3basic-promGLI3-300-luc construct in (-) orientation. No mutation.	This work

	promGLI3-300-luc	orientation. No mutation.	
33	Exon I 5' extended	A fragment of human GLI3 (225 bp) containing sequence of exonII, exonI and part of the promoter region adjacent 5' to exonI. Fragment was amplified using RACE pcr on human placenta cDNA and cloned with T-A cloning in the pcr2.1 vector (INVITROGEN). No mutation.	This work
34	p1230promGLI3-300-lacZ	425 bp region containing 358 bp upstream of presumed initiation of human GLI3 as well as exon1 and 22 bp of intron1, cloned in (+) orientation in PstI site of p1230 vector	This work
35	pGL3basic-Hom2mut1-luc	Hom2 region with more information (434 bp) cloned in HindIII site of pGL3basic vector in (+) orientation. A mutation has been created in the 161-163 nt exchanging aaa with tcc, creating a BamHI site.	This work
36	pGL3basic-Hom2mut2-luc	Hom2 region with more information (434 bp) cloned in HindIII site of pGL3basic vector in (+) orientation. A mutation has been created in the 180-185 nt exchanging aaggaa with ggatcc, creating a BamHI site.	This work
37	pGL3basic-Hom2mut3-luc	Hom2 region with more information (434 bp) cloned in HindIII site of pGL3basic vector in (+) orientation. A mutation has been created in the 215-219 nt exchanging aggaa with gatcc, creating a BamHI site.	This work
38	pGL3basic-promGLI3-300-mut1-luc	425bp region containing 358 bp upstream of presumed initiation of human GLI3 as well as exon1 and 22 bp of intron1, cloned in (+) orientation in XhoI site of pGL3basic vector. A mutation has been created in the 36-40 nt exchanging gacatc with aagctt, creating a HindIII site.	This work
39	pGL3basic-promGLI3-300-mut4-luc	425 bp region containing 358 bp upstream of presumed initiation of human GLI3 as well as exon1 and 22 bp of intron1, cloned in (+) orientation in XhoI site of pGL3basic vector. A mutation has been created in the 145-153 nt exchanging cggggaggg with aagctt, creating a HindIII site.	This work
40	pGL3basic-promGLI3trunc-luc	299 bp region containing 232 bp upstream of presumed initiation of human GLI3 as well as exon1, cloned in (+) orientation in XhoI site of pGL3basic vector.	This work
41	pGL3basic-promGLI3-300-mut2-luc	425 bp region containing 358 bp upstream of presumed initiation of human GLI3 as well as exon1, cloned in (+) orientation in XhoI site of pGL3basic vector. A mutation has been created in the 72-73 nt exchanging gg with aa.	This work
42	pGL3basic-	425 bp region containing 358 bp upstream of presumed	This work

	promGLI3-300-mut3-luc	initiation of human GLI3 as well as exon1, cloned in (+) orientation in XhoI site of pGL3basic vector. A mutation has been created in the 95-96 nt exchanging gg with cc.	
43	pGL3basic-promGLI3-300-mut5-luc	425 bp region containing 358 bp upstream of presumed initiation of human GLI3 as well as exon1 and 22 bp of intron1, cloned in (+) orientation in XhoI site of pGL3basic vector. A mutation has been created in the 118 nt exchanging a with c, creating a HindIII site.	This work
44	PGL3basic-Inrmut	425 bp region containing 358 bp upstream of presumed initiation of human GLI3 as well as exon1 and 22 bp of intron1, cloned in (+) orientation in XhoI site of pGL3basic vector. A mutation has been created in the 343 – 344 nt, exchanging each of the two nt of the int site predicted with primer extension in skeletal muscle cDNA bank with a t.	This work
45	PCR2.1-TOPO	Vector for T-A cloning	Invitrogen
46	pGL3basic	Firefly Luciferase Reporter vector	Promega
47	pGL3prom	Firefly Luciferase Reporter vector	Promega
48	pRLSV40	Renilla Luciferase Reporter vector	Promega
49	pGBKT7		BD

The vectors used for creating the constructs to be used for the cell assays or to be injected in the mice are shown in figure 4-1 and figure 4-2.

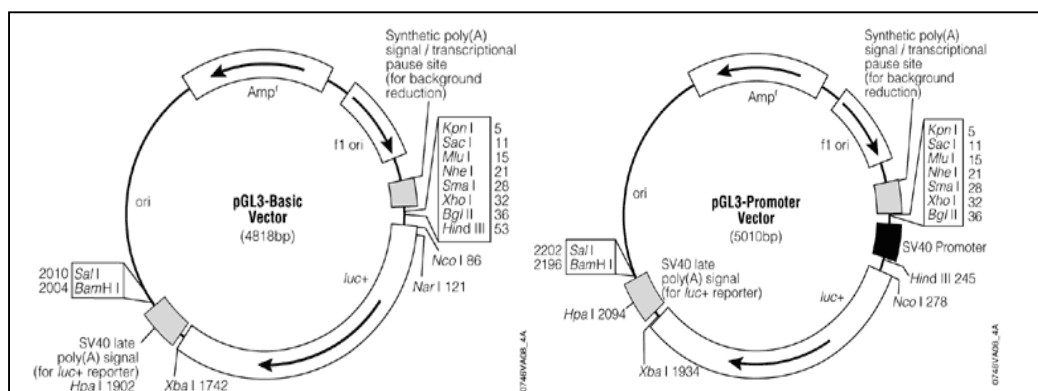


Figure 4-1: Vectors for transient cell transfection assays

pGL3basic is the vector in which all the fragments have been cloned to be tested for their ability in regulating the luciferase reporter gene expression. pGL3promoter is the same vector as pGL3basic containing the SV40 promoter to drive expression of the luciferase reporter gene. It is used as a positive control and its activity is used as a measure for the activity of the cloned fragments in the pGL3basic vector.

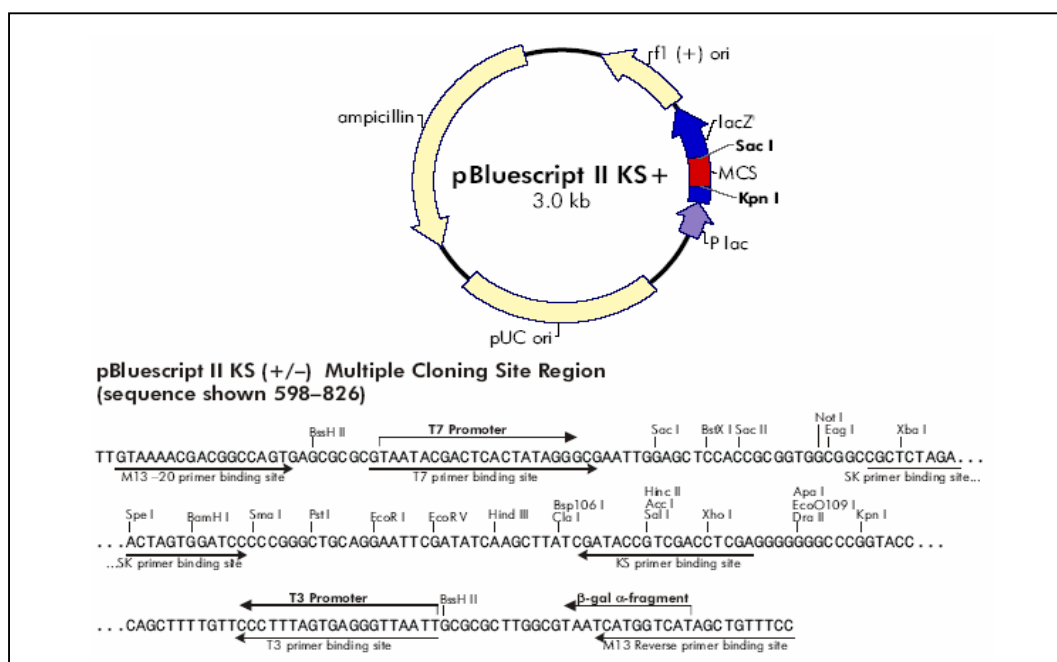


Figure 4-2: p1230 vector

The pBluescript II KS + was used to create the p1230 construct with an 80 bp region of the b-globin promoter with TATA box cloned next to the lacZ gene with an ATG and SV40 polyA signal. This combination is contained on an approximately 3,7 kb Sma fragment which is cloned into the Bluescript Sma site. The b-globin promoter is cloned just downstream of the T3 promoter. The vector shows very little ectopic expression and is preferred for testing enhancer elements in transgenic work.

4.1.10 Organisms

4.1.10.1 Chemically competent bacteria

TOP10

Genotype: F- *mcrA* Δ (*mrr-hsdRMS-mcrBC*) ϕ 80*lacZ* Δ M15 Δ *lacX74* *recA1* *araD139* Δ (*araleu*) 7697 *galU* *galK* *rpsL* (StrR) *endA1* *nupG* (INVITROGEN).

4.1.10.2 Cell lines

NCI-H441

Origin: Claracells, not small cellular lung cancer (NSCLC)
Source: American Type Culture Collection (ATCC HTB-174)
Growth Medium: RPMI 1640, 4% FCS
GLI3 mRNA undetected in Northern Blot (Jacobsen 1996)

NCI-H661

Origin: Not small cellular lung cancer (NSCLC)
Source: American Type Culture Collection (ATCC HTB-183)
Growth Medium: RPMI 1640, 10% FCS
GLI3 mRNA detected in Northern Blot (Jacobsen 1996)

U2OS

Origin: Osteosarcoma cells, epithelial bone cells
Source: American Type Culture Collection (ATCC HTB-96)
Growth Medium: RPMI 1640, 10% FCS
GLI3 mRNA detected in Northern Blot (Topp 1997)

4.1.11 Data banks and online software

Databank/On line software	Application	Internet address
Pubmed	Literature search	http://www.ncbi.nlm.nih.gov/entrez
OMIM (Online Mendelian Inheritance in Man)	Genetic disorders catalogue	http://www.ncbi.nlm.nih.gov/OMIM
NCBI (National Center for Biotechnology Information)	Genome resource	http://www.ncbi.nlm.nih.gov/
UCSC Genome Bioinformatics	Genome data bank	http://genome.ucsc.edu/
TESS	Transcription factor binding site search	http://www.cbil.upenn.edu/cgi-bin/tess/tess
Vista	DNA alignment	http://genome.lbl.gov/vista/index.shtml
ClustalW	Protein alignment	http://www.ebi.ac.uk/clustalw/
Expasy Translate Tool	DNA to protein translation	http://www.expasy.org/tools/dna.html
HGMP Resource Centre	Bioinformatics resource centre	http://www.hgmp.mrc.ac.uk/
Elektronische Zeitschriften Bibliothek	Literature source	http://rzblx1.uni-regensburg.de/ezeit

4.2 METHODS

4.2.1 Single strand conformation analysis (SSCA)

4.2.1.1 PCR amplification of fragments of the GLI3 coding region

PCR was used for amplifying DNA fragments in tandem of the coding region of the GLI3 gene using different pairs of primers. The genomic DNA-template was extracted from blood of patients with GREIG syndrome. PuReTaq Ready To Go PCR beads of Amersham Biosciences were used for the PCR. Each bead contains stabilizers, BSA, dNTPs, 2.5 units of puReTaq DNA polymerase and reaction buffer.

For every PCR one PCR bead, 50 ng of DNA template, 50 ng of each primer, eventually 1 µl formamid and aqua dest water (final volume 25 µl) were used.

4.2.1.2 Non-denaturing polyacrylamide gel for mutation analysis

For the mutation analysis through SSCA, vertical 12% polyacrylamide gels were used. The size of the gels was 16 cm x 18 cm x 0.075 cm. 7.5 ml of 40% acrylamid, 3 ml of 2% bisacrylamid, 2.5 ml of 5x TBE, 235 µl of 10% APS and 15 µl of temed were mixed in a final volume of 25 ml by adding sterile H₂O.

The glass plates were washed thoroughly with aqua dest and isopropanol. After the cast of the gel, combs were used to create 1.2 cm x 2.5 mm x 0.75 mm wells. The gel was let to polymerize for 3hr and in the meantime the DNA probes were prepared (3.5 µl PCR product, 3.5 µl aqua dest and 8 µl loading buffer) and denatured at 95°C for 5'. After removing the comb, the wells were washed with 0.5x TBE and the probes were loaded with a Hamilton pipette. All probes were let to run for 10' at 200 V and 90' at 600 V in 10°C and 20°C.

4.2.1.3 Silver staining of the native polyacrylamide gels

The visualization of the bands occurred through silver staining of the polyacrylamide gel, which was washed for 5' in 1 l of 10% EtOH and 0,5% acetic acid and oxidized for 10' in 1 l of 1% HNO₃. The gel was washed with Milli Q H₂O, stained for 20' with 1 l of 12 mM AgNO₃ and washed again with Milli Q H₂O. The gel was developed with 2 l of 280 mM Na₂CO₃ and 0,019% Formaldehyd and stabilized for 5' in 1 l of 10% acetic acid and approximately 10 ml glycerin. Finally, the gels were wrapped in plastic foils, stretched in plastic frames and allowed to dry overnight.

4.2.1.4 Isolation of DNA from polyacrylamide gels

The extra bands that appeared because of a change in the DNA sequence were cut out and soaked in 100 µl of HPLC-H₂O for 5-6 hours at 4°C and overnight at 37°C. 3-5 µl of the eluate were used for reamplifying the mutated fragment through PCR. The PCR product was purified using the QIAquick PCR purification kit.

4.2.1.5 Sequencing PCR

The sequencing PCR reactions were done by using the DYEnamic ET Terminator Cycle Sequencing Kit (Amersham), containing the labeled dNTPs. 40 ng

of DNA was amplified with 3.2 pmol primer and 4 µl of the sequencing reagent premix in a final volume of 10 µl.

4.2.1.6 Sequencing PCR precipitation

To precipitate the sequencing PCR product, 10 µl of water and 2 µl of sodium acetate/EDTA buffer were added to each tube. Then, 80 µl of 95% EtOH were added in each reaction which were incubated for 20 min in RT in darkness. The reactions were centrifuged for 15' at 14000 rpm and the pellets were washed with 300 µl of 70% EtOH. After a second centrifugation at 14000 rpm for 10' the supernatant was removed and the pellets were air dried.

4.2.1.7 Sequencing electrophoresis and data analysis

The precipitated pellets were resuspended in 4 µl of formamide loading dye and analyzed on a 5% denaturing polyacrylamide gel in an ABI 377 automated sequencer. The sequencing data was obtained from the ABI 377 automated sequencer by Sequence Analysis Software ver 3.4.1 and was analyzed by Sequencher software ver 4.2.

4.2.2 Reverse transcription PCR (RT-PCR)

RT-PCR generates cDNA fragments from RNA templates. In an autoclaved tube 5 µl total-RNA (1 ng/µl) (Clontech) was mixed with 1 µl (500 ng) Oligo(dT)¹²⁻¹⁸ primer and 5 µl RNase free H₂O in a total volume of 10 µl. To avoid the possible secondary structure of the RNA, which might interfere with the synthesis, the mixture was heated at 70°C for 10', and then quickly chilled on ice. After a brief centrifugation, 4 µl of 5x 1st Strand Buffer, 2 µl of 10 mM dNTPmix, 2 µl of 100 mM DTT and 1 µl of PowerScript Reverse Transcriptase (Clontech) were added in the reaction mix. The content of the tube was mixed gently and incubated at 42°C for 90' and then at 70°C for 15'.

4.2.3 Polymerase chain reaction - PCR

The amplification cycles were performed either in a PE2400 or PE9600 automatic thermocycler. The PCR reaction contained in general 10 ng of DNA, 0.2 pmol/µl of each primer, 0.2 µM dNTPs, 1x PCR buffer and 2.5 u of *Taq* DNA Polymerase in a final volume of 25 µl.

4.2.4 Agarose gel electrophoresis of DNA

Usually, 0.5 – 1 g of agarose was added in 50 ml 1.0x TAE buffer, and boiled in the microwave to dissolve the agarose, then cooled down to about 60°C before adding ethidium bromide (10 mg/ml). This 1 - 2% agarose gel was poured into a horizontal gel chamber.

4.2.5 5' RACE PCR

RACE PCR (Rapid Amplification of cDNA Ends) is a technique for amplifying the 5', the 3' end or the whole sequence of a specific cDNA fragment. 5' RACE was used to amplify further the already published cDNA of human GLI2 and GLI3 towards the 5' direction. As a starting template, Marathon-Ready cDNA (BD Biosciences) of human placenta was used. The characteristic of this cDNA library is that each cDNA fragment contains an adaptor sequence at its 5' end. Two different adaptor specific primers, AP1 and the nested AP2 primer, can anneal on this sequence. PCR amplification of this domain is feasible, by using a primer that anneals on the 3' part of the cDNA fragment. The primers must be 23-28 nt long, 50-70 % GC and $T_m > 70^{\circ}\text{C}$. The ones that were used in this work can be found in the primer list for RACE-PCR in the Materials section. The PCR was carried out according to the 5'RACE PCR protocol of BD Biosciences. 5 μl of Marathon-Ready cDNA were mixed with 0.2 μM AP1 primer and 0.2 μM gene specific antisense primer, 1x Advantage 2 Polymerase Mix in a final volume of 50 μl containing 1x cDNA PCR Reaction Buffer. The cycler conditions (for a PE9600 automatic cycler) are given below:

98 ⁰ C : 30''	
98 ⁰ C : 5''	} x 5cycles
72 ⁰ C : 4'	
98 ⁰ C : 5''	} x 5cycles
70 ⁰ C : 4'	
98 ⁰ C : 5''	} x 27cycles
68 ⁰ C : 4'	

The PCR product was diluted 25 times and 5 μl of it were mixed with 0.2 μM AP2 primer and 0.2 μM nested gene specific antisense primer, 1x Advantage 2 Polymerase

Mix in a final volume of 50 µl containing 1x cDNA PCR Reaction Buffer. The cycler conditions (for a PE9600 automatic cycler) for this nested PCR are given below:

98⁰C : 30''
98⁰C : 5'' }
72⁰C : 4' } x 5cycles
98⁰C : 5'' }
70⁰C : 4' } x 5cycles
98⁰C : 5'' }
68⁰C : 4' } x 17cycles

4.2.6 Plasmid isolation using QIAGEN QIAprep spin miniprep kit

To obtain plasmid of high purity for DNA sequencing, the target clone was inoculated in 5 ml of LB liquid medium containing 50 µg/ml ampicillin at 37°C while shaking at 225 rpm for 12-16 hours. The culture was then centrifuged at 3000 rpm for 10' at 4⁰C and the pelleted bacterial cells were resuspended in 250 µl Buffer P1 and transferred to a microcentrifuge tube. Then 250 µl Buffer P2 were added, and the tube was gently inverted 4–6 times to mix. After 2'-5' 350 µl Buffer P3 was added, the samples were mixed again by inversion and centrifuged for 10' at 13,000 rpm (~17,900 x g) in a table-top microcentrifuge. The supernatants were applied to the QIAprep spin column by decanting or pipetting and centrifuged for 30''–60''. The flow-through was discarded and the QIAprep spin column was washed by adding 0.75 ml Buffer PE and centrifuging for 30''–60''. The flow-through was removed and the samples were centrifuged for an additional 1 min to remove residual wash buffer. Finally the QIAprep column was placed in a clean 1.5 ml microcentrifuge tube. DNA was eluted by adding 50 µl or water to the center of each QIAprep spin column, letting stand for 1', and centrifuging for 60''.

4.2.7 Medium-scale preparation of plasmid DNA (Qiagen Midi- Kit)

A single clone was inoculated in 2 ml LB medium with appropriate antibiotic as a pre-culture for 8 hrs in 37°C shaker. This starter culture was diluted 1/500 – 1/1000 into 100 ml of selective LB medium and incubated 12-16 hrs at 37°C with shaking (300 rpm). The saturated culture was centrifuged for 15 min at 6000 x g at 4⁰C. The pellet was resuspended in 4 ml of solution P1 and cells were lysed with 4 ml P2 and

neutralised with 4 ml P3. The precipitated solution was incubated on ice for 15' and centrifuged at 20.000 x g for 30' and the supernatant once more for 15' at 4°C. Meanwhile, the column (Qiagen-tip) provided with the Midi preparation kit was equilibrated with 4 ml of QBT solution. After centrifugation the lysate was poured into this equilibrated column thus allowing the DNA to bind with the resin present in the bed of the column. The column was then washed twice with 10 ml of solution QC. Finally, the DNA was eluted with 5 ml of QF solution. To precipitate the DNA, 3.5 ml of room temperature isopropanol was added and mixed thoroughly and centrifuged at 14000 x g for 30' at 4°C. The DNA pellet was washed with 2 ml 70% room temperature ethanol and dissolved in 60 µl of TE or H₂O.

4.2.8 Large-scale preparation of plasmid DNA (Qiagen Maxi- Kit)

A single clone was inoculated in 2 ml LB medium with appropriate antibiotic as a pre-culture for 8 hours in 37°C shaker. In 100 ml LB medium with appropriate antibiotic this pre-culture was added in a dilution of 1/100 fold and incubated overnight at 37°C with shaking. The saturated culture was centrifuged for 15 min at 6000 g at 4°C. The pellet was resuspended in 10 ml of solution P1 and cells were lysed with 10 ml P2 and neutralised with 10 ml P3. The precipitated solution was incubated on ice for 20' and then centrifuged at 10,000 rpm for 30' at 4°C. The supernatant was transferred in a fresh tube and centrifuged again at 10,000 rpm for 15' at 4°C. Meanwhile, the column (Qiagen-tip) that was provided with the Maxi preparation kit was equilibrated with 10 ml of QBT solution. After centrifugation the supernatant was poured into the equilibrated column thus allowing the DNA to bind with the resin present in the bed of the column. The column was then washed twice with 30 ml of solution QC. Finally, the DNA was eluted with 15 ml of QF solution. To precipitate the DNA, 10.5 ml of isopropanol was added and mixed thoroughly and centrifuged at 14000 g for 30' at 4°C. The DNA pellet was washed with 70% ethanol and dissolved in 200 µl of TE or H₂O.

4.2.9 RNEASY Mini Protocol for total RNA isolation from cells

All the steps were carried out in room temperature and in RNase free conditions. First, 10⁶ cells in PBS were pelleted by centrifugation at 1500 rpm for 5'. The supernatant was removed and the pellet was diluted in 350 µl buffer RLT (10µl β-mercaptoethanol were added in 1ml RLT buffer). The mixture was homogenized by

pipetting 5-10 times with a 20-gauge needle. Next, the RNA was precipitated with equal volume of room temperature 70 % EtOH and the sample was applied on the column. After centrifugation for 15'' at 10000 rpm the flowthrough was discarded and the column was washed with 700 µl RW1 buffer (15'' – 10000 rpm) and then, after removing the flowthrough, it was washed again twice with 500 µl RPE buffer (first 15'' – 10000 rpm, then 2' – 10000 rpm). The flowthrough was discarded, and the column was placed in a clean 1.5 ml Eppendorf tube. 30 µl of RNase-free H₂O were added on the column, which was then centrifuged for 1' at 10000 rpm.

The concentration of the RNA was determined with a photometer at 260 nm.

4.2.10 QIAquick Gel Extraction method

This method is designed to extract and purify DNA of 70 bp to 10 kilobase pair (kb) in length from agarose gels. Up to 400 mg agarose can be processed per spin column. The principle of this method depends on selective binding of DNA to uniquely designed silica-gel membrane. The excised DNA fragment in an agarose slice was weighed, and three volumes of QG buffer were added and incubated at 50°C for 10-15 min. After the gel slice was completely dissolved, 1 gel volume isopropanol was added to the sample, mixed and then applied to a QIAquick column and centrifuged for 1'. The flow through was discarded and the column was washed with 750 µl of PE buffer. After drying the column it was placed into a fresh microcentrifuge tube. To elute DNA, 35-50 µl of EB buffer or water was applied to the centre of the QIAquick membrane and centrifuged for 1'.

4.2.11 Restriction enzyme digestion of DNA

Restriction enzyme digestions were performed by incubating double-stranded DNA with an appropriate amount of restriction enzyme in its respective buffer as recommended by the supplier, and at the optimal temperature for that specific enzyme. Standard digestions included 2-10 u enzyme per microgram of DNA. These reactions were usually incubated for 1-3 hrs to ensure complete digestion at the optimal temperature for enzyme activity, which was typically 37°C.

4.2.12 Klenow treatment

The isolated DNA fragment was incubated with 1x Buffer O⁺ (MBI Fermentas), 0.05 mM dNTPs and 1.5 u of Klenow enzyme in a final volume of 20 µl.

4.2.13 CIP treatment of digested vector

10 u of CIP (NEB) were added to the digested vector, and the reaction took place at 37°C for 1 hour.

4.2.14 Ligation (Invitrogen)

To calculate the necessary amount of insert and vector, which should be used in the ligation reaction, the following formula was used:

$$\text{Insert (ng)} = 5 \times \text{vector (ng)} \times \text{insert (bp)} / \text{vector (bp)}$$

The necessary amount of vector and insert were mixed with 5 u T4 DNA ligase in a reaction volume of 10 µl containing 1x ligation buffer. The ligation reaction was carried out at 14°C overnight.

4.2.15 TOPO TA Cloning

4 µl of PCR product were incubated with 1 µl of salt solution and 1 µl of pCR2.1-TOPO vector at RT for 5'. If necessary, prior to the cloning, the PCR product was incubated with 5 u of taq polymerase and 0.3 mM dNTPs at 72°C for the addition of the A residues.

4.2.16 Chemical transformation of TOP 10 E.coli competent cells

In order to obtain plasmid DNA in large amounts for further experiments, the ligation product was transformed into an aliquot of TOP 10 *E.coli* competent bacterial cells. The competent cells were removed from a -80°C freezer and thawed on ice. Next, 6 µl of the ligation product were added to the aliquot of the cells. The tube was gently tapped to mix and then incubated on ice for 30'. To heat shock the cells after the incubation, the sample was placed in a 42°C water bath for 30'', and then placed on ice for two minutes. 250 µl of SOC medium was added, and the cells were incubated for 1 hour at 37°C while shaking at 225 rpm. 50 µl of the culture was spread out on LB agar growth plates containing 50 µg/ml ampicillin. The plate was incubated in a bacterial cell incubator at 37°C overnight. Transformed bacteria form a single colony growing on the appropriate antibiotic selective medium. The growth plates were stored at 4°C.

4.2.17 Liquid culture

Single bacterial clones that grew on the LB-Agar Amp growth plate were picked with plastic pipet tips and incubated in 3 ml of LB liquid medium containing 50 µg/ml ampicillin at 37°C while shaking at 225 rpm for 12-16 hours.

4.2.18 Small-scale isolation of plasmid DNA

1,5 ml of the grown cultures in the liquid medium were transferred in 1.5 ml Eppendorf tubes and sedimented by centrifugation for 30'' at 14.000 rpm. The pellet was re-suspended in 250 µl of solution P1. The bacterial cells were lysed with 250 µl of P2 solution and then neutralised with 250 µl of solution P3, which precipitates the proteins. The precipitated solution was incubated on ice for 5', and centrifuged at 14000 rpm at 4°C. The supernatant was transferred into a new tube, and 0,8 ml of ice cold 100% isopropanol was added to precipitate the DNA. It was then centrifuged for 15', and finally the pellet was washed with 70% ethanol. After air-drying the plasmid-DNA pellets were dissolved in 50 µl of dH₂O.

4.2.19 Plasmid DNA sequencing

In order to control clones for the presence of any point mutations generated during the process of amplification, appropriate DNA preparations were sequenced. Plasmids were prepared using the QIA plasmid prep kit as described above. The purity and concentration were analysed spectroscopically. For sequencing a plasmid insert, plasmid concentrations of 100 ng/kb were used in a reaction mix of 10 µl which contained 0.4 µl of 10 µM primer and 2 µl of termination mix (Big Dye Terminator v1.1 Cycle Sequencing Kit). A standardized PCR reaction program was used with an initial denaturation at 95°C for 5', 24 cycles of denaturation at 94°C for 20'', primer annealing at T_m of maximum 60°C for 15'', extension at 60°C for 45'', and terminated with 60°C for 4' to facilitate the completion of extension reaction. After PCR completion, the products were precipitated with 75% isopropanol for 1hour in darkness and centrifuged for 20' at 14000 rpm. The pellet was washed with 70% ethanol. The sample was dried in a thermoblock at 90°C for 2' and resuspended in 15 µl High Dye Formamide before subjecting to analysis by the ABI 310 sequence analyser.

4.2.20 Cell techniques**4.2.20.1 Culture of eukaryotic cells**

Handling and propagation of all cell lines were performed in a cell and tissue culture laminar-air flow under sterile conditions. All solutions were stored at 4°C and warmed up to 37°C in a water bath before using. All solutions were only opened under sterile conditions in a laminar flow. The adherent H661 (lung cancer) cells were grown in RPMI medium, 10% FCS, 1% non essential amino acids, 2% Penicillin/Streptomycin and 1% L-glutamine cultivated in the incubator at 37° C, 5% CO₂ and 88% moisture.

4.2.20.2 Freezing cultured eukaryotic cells

To freeze cells for long term storage, they were harvested (at least one T 75 flask of 80% confluent cells) and centrifuged at 1,000 g for 5'. The media was then aspirated, the cells were re-suspended in 1 ml of cell cryoprotective medium and frozen in cryovials at -80°C.

4.2.20.3 Splitting the eukaryotic cell cultures

All cell lines were grown in a 37°C incubator and split in certain ratios depending on the stage of confluence and the proliferation rate of each cell line. The amount of medium added to the flask was dependent on its size. 10 ml were added into medium sized flasks (T75). The medium was changed every second day. To split cells, the medium was completely aspirated from the flask. In order to detach adhering cells PBS/EDTA (half the amount of culture medium) was added to a culture flask. The cells detached after about 5' at 37°C with occasional gentle tapping, and after removal of PBS/EDTA by centrifugation an aliquot of the resuspended cells was transferred into a flask containing fresh medium.

4.2.20.4 Maintenance of human cell lines

The cell line aliquots were stored in 1 ml cryoprotective medium at -80°C. To grow up a cell line, the frozen cells were thawed quickly in a 37°C water-bath then transferred into a 15 ml tube and mixed with 10 ml of medium. The suspended cells were centrifuged at 1,000 g for 5' in order to isolate a cell pellet free of DMSO residues from the freezing medium. After removing the supernatant, the pellet was

resuspended in 10 ml medium and transferred into a cell culture flask. The cells were stored in a 5 % CO₂, 37°C incubator.

4.2.20.5 Cell counting

The cells were detached using 5 ml of PBS/EDTA and incubation for 5' at 37⁰C, centrifuged for 5' at 1200 rpm in a 15 ml Falcon tube. The supernatant was removed and the cells were resuspended in 5 ml PBS. This washing step was repeated once more, and finally one aliquot of the resuspended cells was used for counting. The counting of the cell density was performed in a Neubauer cell counting chamber.

4.2.20.6 Transient transfection of adherent cells

Transient transfection is a technique used for inserting DNA constructs in cell cultures usually to express a specific protein or to test the potential of promoters or other cis regulatory elements to drive expression of reporter genes in specific cell lines.

The day before transfection, in each well of a 12-well plate a sufficient number of cells were plated (2×10^5 H661, 1.5×10^5 U2OS cells or 3.5×10^5 H441 cells) in 2ml suitable growth medium to reach on the next day, when the transfection was to be performed, 60-80% confluency. The day of transfection, 500 ng of total DNA in 5 μ l TE buffer pH 7-8 (200 ng test construct, 100 ng Renilla construct, pRLSV40, 200 ng stuffer DNA) (table 4-1) were mixed with the DNA-condensation buffer, Buffer EC, to a total volume of 75 μ l. 4 μ l of enhancer were added, and the mixture was vortexed for 1''. The mix was incubated at room temperature (15–25°C) for 2–5' and then spun down for a few seconds to remove drops from the top of the tube. Finally, 6 μ l effectene transfection reagent were added to the DNA-enhancer mixture, mixed by pipetting and incubated for 5–10' at room temperature (15–25°C) to allow transfection-complex formation. In the meantime, the cells were washed with 4 ml PBS and provided with 0.8 ml fresh growth medium (can contain serum and antibiotics). 0.4 ml growth medium (can contain serum and antibiotics) were added to the tubes containing the transfection complexes, mixed by pipetting and immediately the transfection complexes were added drop-wise onto the cells in the 12 well plate. The plates were gently swirled to ensure uniform distribution of the transfection complexes and incubated under normal growth conditions for 48 hr for transfection to occur and for the expression of the transfected gene.

Each of the constructs had been purified with QIAquick mini/midi prep kit and brought to a concentration of 100 ng/μl. The DNA mixtures used for the transient transfections are given in the following table:

	pGL3basic	pGL3prom	Sample	pRL-SV40	pGBKT ₇
Negative control	2 μl (200 ng)	-	-	1 μl (100 ng)	2 μl (200 ng)
Postitive control	-	2 μl (200 ng)	-	1 μl (100 ng)	2 μl (200 ng)
Sample	-	-	2 μl (200 ng)	1 μl (100 ng)	2 μl (200 ng)

Table 4-1: Concentrations of each construct necessary for transient cell transfection

On the next day, 1.5 ml of complete medium was added to each well and culture was continued at 37°C.

For each construct 3-6 transfection experiments were performed in parallel.

4.2.20.7 Protocol for dual luciferase assay on a 12 well plate

The transfected adherent cells in each well were washed with 1ml PBS and then treated with 250 μl of 1x PLB (Passive Lysis Buffer). The culture plates were placed on an orbital shaker with gentle shaking to ensure complete and even coverage of the cell monolayer with 1X PLB for 15'. Then, the lysate was transferred into a 1.5 ml Eppendorf tube and centrifuged for 30'' at top speed in a refrigerated microcentrifuge. 20 μl of the cleared lysate were transferred into a clean polypropylene tube. 1.2 ml of LAR II buffer and 1.2 ml of Stop and Glo buffer were mixed with 200 μl water, each. LAR II and Stop and Glo are necessary for the light emission of the firefly and the renilla luciferase respectively. To the Stop and Glo buffer, 12 μl of 100X Stop & Glo® substrate were added. The luciferase activity was measured with the AutoLumat LB 953 luminometer.

4.2.21 Primer extension assay

The primer extension assay is a useful method for determining the transcription start site of genes. A primer 22-30 nt long marked with IRD-700 on its 5' end is used to reverse transcribe mRNA to cDNA. The primer must be specific for the gene of interest, must anneal not further than 200 – 300 nt of the presumed transcription

initiation site and must have an extension direction towards the 5' end of the gene. The resulting product is precipitated like a sequencing PCR reaction product. A genomic fragment, which contains at least the sequence where the marked primer anneals and the presumed initiation site is undergoing a sequencing PCR reaction with the same primer used for the primer extension and precipitated. Both the cDNA that resulted from the marked primer and the genomic sequence are let to run parallel in a sequencing gel. The marked cDNA should give a signal right next to the nucleotide of the genomic sequence, where the cDNA sequence stops. By this way, the transcription initiation site of a gene is determined.

4.2.21.1 Primer extension reaction

For the primer extension, total RNA was used that expresses the gene of interest. 5 µg of total RNA in RNase free H₂O were mixed with 2 pmol of 5' IRD700 modified primer and incubated for 10' at 70°C in a final volume of 11 µl. Then, to the reaction were added 1 mM dNTP mix, 2 µl of DTT, 1 u of Powerscript RT and first strand buffer in a final concentration of 1 µl. The mixture was incubated for 90' at 42°C and for 15' at 70°C.

4.2.21.2 Precipitation of primer extension product

The product of the primer extension (20 µl) was diluted in 175 µl of TES buffer (10 mM Tris-Cl, pH 7.5; 1 mM EDTA, pH 7.5; 0.1% (w/v) SDS). To remove all the proteins from the mixture, it was mixed with 50 µl phenol, 48 µl chloroform and 2 µl of isoamylalcohol and centrifuged for 5' at 13000 rpm. The supernatant was transferred into a clean Eppendorf tube, and 1 µl of glycogen (10 µg/µl) was added to enhance the precipitation together with 600 µl of ice cold absolute ethanol. After at least 1 hr incubation in -80°C the sample was centrifuged for 15' at 13000 rpm and the supernatant was removed. After the pellet was vacuum dried, it was diluted in 6 µl of HPLC water and stored at 4°C.

4.2.22 Transgenic mice

Transgenic mice are produced by the microinjection of DNA constructs into the pronuclei of 0.5 dpc (days post coitum) one-cell fertilized eggs and the injected eggs are transferred to the oviducts of 0.5 dpc pseudopregnant foster mothers. The DNA construct integrates randomly into the genome of the egg and the resulting founder

mice will contain one or more copies of the transgene in every cell. By this way, the integrated construct of cis regulatory element – reporter gene drives expression of the reporter gene in the tissues in which the element specializes.

4.2.22.1 DNA isolation and precipitation prior to injection

The constructs that contained the fragments to be injected were double digested with Not/XhoI (NEB). The digestion product of each reaction was run in a 1% agarose gel and the DNA band corresponding to the insert was extracted with the QIAquick Gel Extraction method. The isolated, linearized DNA fragment was then precipitated with 1/10 volume of sodium acetate, double volume of 100% EtOH and incubated for 30' at -80°C. The mixture was centrifuged at 14.000 rpm for 15', the supernatant was removed and the pellet was washed with 200µl of 70% EtOH.

4.2.22.2 Mice injection

The isolated DNA fragments were diluted in 10 mM Tris, pH 7.5, 0.1 mM EDTA, pH 8.0 buffer in a final concentration of 1-3 µg/ml. The amount of DNA injected cannot be determined absolutely, but it is estimated that 1-2 pl are microinjected into each male pronuclei of fertilized eggs. These eggs were then transferred in the oviduct of pseudopregnant female mice, which give birth to transgenic mice. These mice have integrated the DNA fragment into their genomic DNA of almost all of their cells. However, because the integration happens randomly, the same construct can cause different staining in the different mice, as the transcription of the reporter gene depends on the region of the chromosomes in which it is integrated. Therefore, if interesting staining is observed, it is preferred to create a stable line of mice containing the fragment inserted in the same chromosomal region.

4.2.22.3 Analysis of embryos

The time of gestation was calculated taking noon of the day of detection of a vaginal plug as embryonic day 0.5 (E 0.5). Expression of the transgene was detected by staining staged embryos overnight in X-gal buffer. Embryos were dissected free of extraembryonic membranes (which were retained for genotype analysis) then fixed in 4% paraformaldehyde at 4°C for 30' to 2 hours, depending on their developmental stage, washed in detergent wash (2 mM MgCl₂, 0.05% BSA, 0.1% sodium deoxycholate, 0.02% NP-40 in 0.1 M sodium phosphate buffer, pH 7.3), and stained

overnight in detergent wash containing 5 mM $K_3Fe(CN)_6$, 5 mM $K_4Fe(CN)_6$ and 0.085% NaCl with 0.1% Xgal at 37°C. The staining reaction was stopped by washing the embryos in detergent wash then in phosphate-buffered saline. The embryos were postfixed overnight in 4% paraformaldehyde at 4°C.

Detergent wash:

2 mM $MgCl_2$

0.05% BSA

0.1% sodium deoxycholate

0.02% NP-40

in 0.1 M sodium phosphate buffer, pH 7.3

Sodium phosphate buffer:

Na_2HPO_4/NaH_2PO_4 mixed in right ratio

Staining:

2 mM $MgCl_2$

0.05% BSA

0.1% sodium deoxycholate

0.02% NP-40

in 0.1 M sodium phosphate buffer, pH 7.3

5 mM $K_3Fe(CN)_6$

5 mM $K_4Fe(CN)_6$

0.085% NaCl

0.1% Xgal (add at last moment)

4.2.22.4 DNA isolation from mice

The yolk sac of each transgenic mouse embryo was incubated in 300 µl of a 50 mM Tris pH 8.0, 5 mM EDTA, 0.5% SDS buffer and 10 µl of 10 mg/ml proteinase K at 55°C overnight. The next day, 100 µl of 6 M NaCl was added and the samples were vortexed and centrifuged at 14,000 rpm for 10''. The supernatant was transferred in a new tube and 300 µl of isopropanol were added. The mixture was centrifuged at 14,000 rpm for 5', the supernatant removed, and the pellet was washed with 250 µl of

75% EtOH, air dried and dissolved in 25-100 μ l water, depending on the size of the starting material.

5 RESULTS

Point mutations, translocations and deletions throughout the *GLI3* gene can cause various autosomal dominant polysyndactyly syndromes summarized as GLI3 morphopathies (Vortkamp et al. 1991; Wild et al. 1997; Kalff-Suske et al. 1999). In patients with these phenotypes, which do not show mutations affecting their GLI3 coding sequences, regulatory sequences might be affected. In order to search GLI3 regulatory sequences for mutations, a collection of cases with potential GLI3 morphopathies without mutations in coding sequences has to be established. Towards this end, the DNA of appropriate patients was analysed for mutations in the GLI3 coding sequences.

GLI3 is a large gene (cDNA 5055 bp) extending over 250 kb on chromosome 7p13. Therefore, mutations were searched in a two-step procedure: First, exonic fragments including intron-exon boundaries were screened for changes in the DNA sequence affecting single strand electrophoretic mobility. Subsequently, fragments with differing electrophoretic behaviour were sequenced to identify the underlying mutation.

5.1 Search for mutations in individuals with putative GLI3 morphopathies

DNA was extracted from the blood of 24 patients with GLI3 morphopathies using standard procedures. The coding exons 2 to 15 of GLI3 including the intron-exon boundaries were amplified by PCR using the primer pairs described above, and subjected to single strand conformation analysis (SSCA) at two different conditions. Exon one was not included as the translation of GLI3 starts from the second exon. The separated strands of PCR-amplified DNA fragments were screened by electrophoresis in a native (non-denaturing) polyacrylamide gel. Point mutations result in a variant three-dimensional structure of the single strand DNA molecules, which influences their electrophoretic behavior.

The sensitivity of SSCA depends on:

- The length of the PCR-fragments (best results are obtained for fragments smaller than 300 bp)
- The GC-content of the sequence
- The kind and the position of the mutation

- The electrophoretic conditions (buffer concentration, temperature)

Large exons were split into conveniently sized, overlapping fragments. Amplicons, in which at least one of the patients showed an electrophoretic band with deviating mobility, were screened in 100 unaffected, unrelated control individuals. Non-polymorphic variants were extracted from the gel and sequenced.

The mutation analysis detected a series of DNA polymorphisms in exons 5, 4, 14, and 15, which had already been described, previously.

Three non-polymorphic missense mutations and one nonsense mutation were observed among the patients.

In a patient with syndactyly in all the limbs, polysyndactyly in the left and right foot, high forehead and bregma, frontal bossing, hypertelorism and broad nasal root a mutation in exon 13 (c.1873 C>T; R625W) was found. In the second case, a patient with hexadactyly featured a mutation in exon 15 (c.3664 C>T; P1222S). Both missense mutations had been previously described in other patients.

Two novel mutations were detected (figure 5-1). One change was identified in the 13th exon (c.2058_2059 GG > AT; R686R, E687X), which represents one silent and one nonsense mutation next to each other. This patient showed a mild syndactyly, postaxial polysyndactyly and broad terminal phalanges in both hands, hypertelorism, broad nasal root, short structure, growth hormone deficiency and hypothalamic hamartoblastoma. The second novel, a missense mutation in exon 15 (c.4675 A>G; I1559V) was detected in a patient reported to show an acrocallosal syndrome.

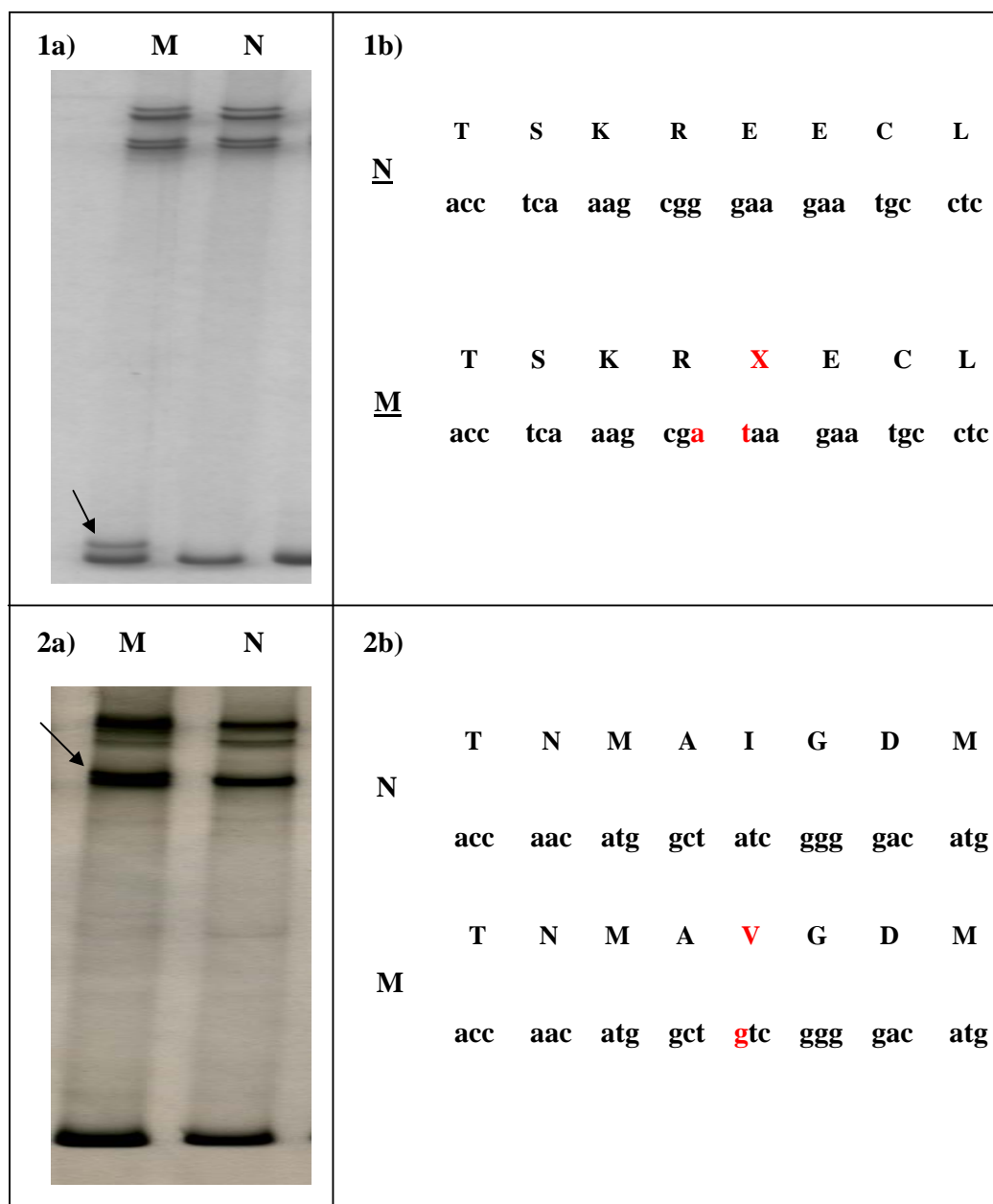


Figure 5-1: Novel mutations detected in the GLI3 gene of patients with GLI3 morphopathies

1) SSCA gel (a) comparing the pattern of exon 13 from the patient (M) and an unaffected control individual (N). The corresponding DNA and protein sequences around the observed mutation are aligned (b).

2) Exon 15 of a Greig syndrome patient (M) and of an unaffected control individual (N). a): SSCA pattern, b): aligned sequences of the mutated region.

Arrows depict the deviating band in the patients. The mutated nucleotides and the resulting amino acids are shown in red.

For the mutation in exon 13, the first nucleotide exchange does not alter the amino acid due to the wobble effect but the neighboring alteration creates a premature stop codon. In the 15th exon, the missense mutation leads to an exchange of isoleucine by valine. The position of the novel mutations is shown in figure 5-2.

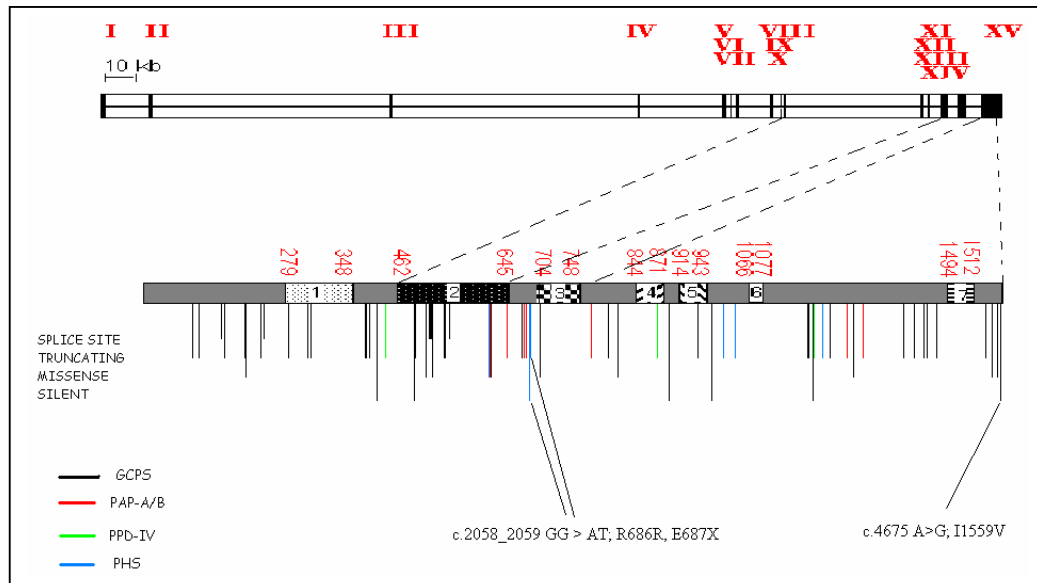


Figure 5-2: Position of the novel mutations on the GLI3 protein

Diagrams of the genomic structure of GLI3 (above) and the protein sequence (below). Exons are indicated with Roman numbers. Above the bar representing the GLI3 protein, amino acids flanking functionally important, conserved domains are shown in red. Below this bar, the position and the type of mutations observed in different GLI3 morphopathies are depicted. The two novel mutations are described individually. (GCPS = Greig cephalopolysyndactyly syndrome; PAP-A/B: polydactylies A and B; PPD-IV: postaxial polydactyly type IV; PHS: Pallister-Hall syndrome).

5.2 Experimental control of the published human GLI2 cDNA sequence

5.2.1 Human GLI2 5'RACE PCR

The overlapping and distinct expression of human GLI2 and GLI3 speaks for the existence of both similar and different transcriptional regulation of these genes. As the promoter region of *GLI2* has not been yet identified, it was decided to find additional upstream sequence by amplifying and sequencing of the 5' region of this gene.

RT-PCR was performed on human placenta and fetal lung total RNA. The resulting cDNA served as a template for *GLI2* specific PCR (primers “*GLI2specF*” & “*GLI2specR*”). Figure 5-3 shows that placenta expresses *GLI2* in abundance, thus this tissue was chosen to proceed with 5’ RACE PCR.

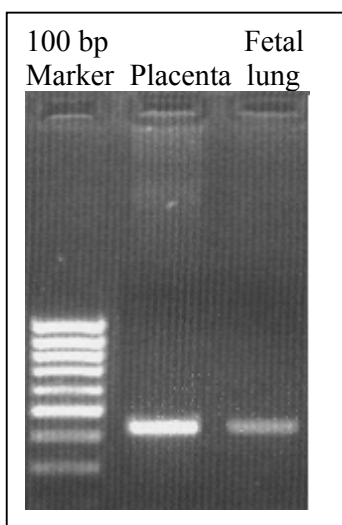


Figure 5-3: PCR with *GLI2* specific primers using cDNA generated by RT PCR on human placenta and fetal lung total RNA

The electrophoretic separation shows that a product of 423 bp is observed in greater abundance from placenta cDNA.

For 5’ RACE PCR, commercially RACE-ready human placenta cDNA was used as a template to investigate the upstream region of *GLI2*. This cDNA library is already ligated with the necessary adaptors to perform the PCR. The single intense PCR fragment obtained by the 5’ RACE PCR using the gene specific primers (primers “*GLI2Race2*”, “*GLI2Race3*”, “*GLI2Race4*”), which anneal in the 2nd exon of *GLI2* according to the published sequence (NM_030381) is shown in figure 5-4.

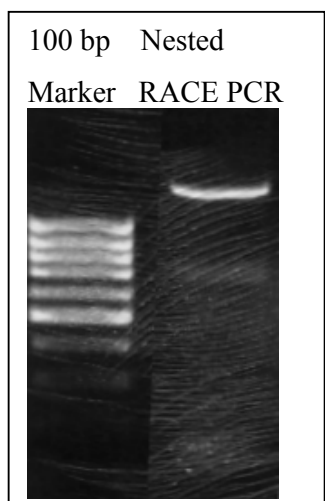


Figure 5-4: Electrophoresis of the RACE PCR product

After the nested PCR reactions one single DNA band with a size larger than 1 kb is observed.

The resulting PCR product was cloned in the TOPO-TA cloning vector that was transformed into TOP 10 *E. coli* competent cells. Positive clones were isolated and sequenced.

The sequence of this amplified region was aligned using the “bl2seq” program from NCBI (<http://www.ncbi.nlm.nih.gov/>) and the “BLAT” program from the USCS database (<http://genome.ucsc.edu/>) with the published GLI2 cDNA and the corresponding genomic sequence. It extended from the GLI2 specific published exon 2, on which the primers for the 5’ RACE PCR annealed, into the published 1st exon and continued for 1195bp further upstream. The sequence ended with the adaptor sequence of the MARATHON RACE READY cDNA LIBRARY, indicating that it did not represent a genomic DNA contamination.

The extended cDNA sequence matches the published genomic sequence of this chromosomal region upstream of the previously determined start of transcription, indicating that it is part of the first exon. However, this region neither shows any similarity with mouse GLI2, nor does it contain an ORF that is compatible with the published exons of human GLI2. All of the possible ORFs end with a stop codon, not allowing the GLI2 protein to be produced. Using the NIX program of the HGMP Resource Centre (<http://www.hgmp.mrc.ac.uk/>), a genomic fragment containing the 1st and the 2nd exon, the 1st intron and 2000 bp upstream of the 1st exon (according to the published sequence [NM 030381]) was searched for potential promoter elements. A CpG island of 208 bp could be identified residing inside the *GLI2* cDNA sequence that was expanded in this work. The only predicted promoter element located upstream of the expanded cDNA is a TATA box 682 bp upstream of the cDNA start. Searching for promoter elements in the upstream region of the murine *Gli2*, 93 bp upstream of its 1st exon ends a 331 bp long CpG island. Another, 452 bp long, CpG island is located 382 bp downstream of the start of this exon. Inside the first CpG island there are predicted two *inr* sites and a TATA-less promoter. These predicted promoter elements of human and mouse GLI2 do not share any similarity. Sequence comparison between the upstream region of human *GLI2* and *GLI3* did not reveal any similarity either, suggesting that the promoter regions of these genes are completely different.

5.3 Identification and functional analysis of the *GLI3* promoter

5.3.1 Determination of transcription start site of human *GLI3*

The 1st exon as well as the promoter region of human *GLI3* is not well characterized. The published sequence of the first exon might not to be complete since there is no transcription start site determined.

Human placenta cDNA was used to amplify the 5' upstream region of *GLI3* by 5'RACE PCR, as this gene is also expressed in this tissue. The reverse primers were designed to anneal to the published 2nd exon of human *GLI3* and directing towards the 5' end of the cDNA. The first primer ("GLI3RT1") anneals 123 nt and the nested primer ("GLI3RT2") anneals 92 nt downstream of the ATG with which the ORF of *GLI3* starts. The PCR product of the nested PCR reaction is shown in figure 5-5.

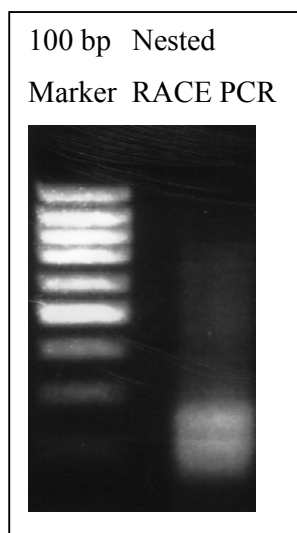


Figure 5-5: Electrophoresis of the products of 5'RACE PCR

A 200-250 bp product is demonstrated.

After the 1st and the nested PCR, a 200-250 bp product was obtained and cloned in a TOPO-TA cloning vector.

A number of clones were isolated and sequenced. The longest transcript using this method was found two times and starts 133 nt upstream of the ATG. The extra sequence that can be found in the placenta cDNA bank is:

CctcgagctgggacccactccgggcttggggacccggcgggcgaggagaagcgggggctcccggcgcggcctgG
AAGGAGCGAGAAAGCAAAAGGGCATT TTTGGTCTGAAGAGAGCTGAAGTA
ATGAGAAGACATCAT**G**

The sequence in black capital letters represents part of the already published sequence of exon 2 (including the ATG (bold black capital letters) with which the translation starts) and exon 1 of human GLI3, and the sequence in black small letters represents the extra sequence that was read in the cDNA library of human placenta with the 5' RACE PCR. Using the NIX analysis program of the HGMP Resource Centre (<http://www.hgmp.mrc.ac.uk/>), a transcription start site was predicted at "CCT" just upstream to the new start sequence of exon 1. This start site is indicated above in bold red letters.

One transcript was found to start 119 nt upstream of the ATG and its sequence is shown with small black letters, upstream of the 1st exon and 2nd exon (counting starts with the A of the ATG (bold black capital letters), with which the translation starts) that are depicted in black capital letters.

CcactccgggcttggggacccggcgggcgaggagaagcgggggctcccggcgcggcctgGAAGGAGCGA
GAAAGCAAAAGGGCATT TTTGGTCTGAAGAGAGCTGAAGTAATGAGAAGA
CATCAT**G**

There is a strong prediction for an initiator site found exactly in this sequence, CCACTCC, which includes two nucleotides more (red bold letters) than this transcript (Jacobsen, 1996)

One transcript was found to start 100 nt, one 46 nt, three transcripts started 44 nt, one 42 nt and one 21 nt upstream of the ATG. One transcript was found to start 25 nt downstream of the ATG.

The whole procedure (RACE PCR, cloning and sequencing) was repeated. The longest transcript started again 136 nt, one transcript 54 nt, four transcripts started 44 nt, one 43 nt, one 42 nt, one 41 nt, one 32 nt and one 20 nt upstream of the ATG with which the ORF of GLI3 starts.

5.3.2 Experimental determination of the transcription start site with the primer extension assay

A primer was designed with a 5' IRD-700 modification directing towards the 5' end of the GLI3 gene. This primer ("Exten") anneals approximately 90bp downstream of the presumed initiator site and was used to synthesize cDNA by RT-PCR from human skeletal muscle total RNA, which also expresses GLI3. This product was used to identify the transcription start site of the GLI3 using the primer extension assay.

The resulting cDNA was precipitated and run in a sequencing gel next to the sequencing reaction of the pGL3basic-promGLI3-300-luc construct which was sequenced with the same primer that was used for the primer extension assay (figure 5-6).

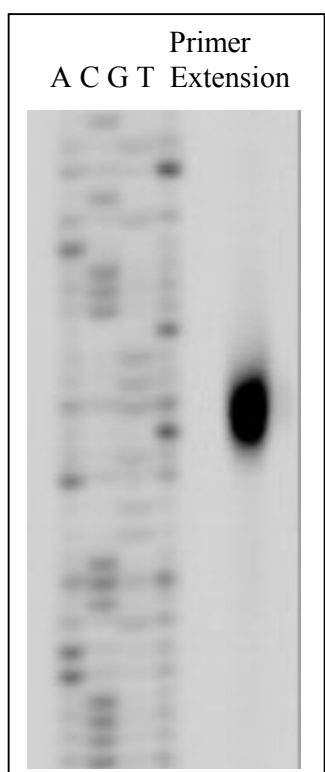


Figure 5-6: Primer extension analysis of the GLI3 transcription start site

The primer extension product next to the sequence of the pGL3basic-promGLI3-300-luc construct. The signal from the primer extension band marks the nucleotide with which the cDNA, thus the mRNA, of GLI3 begins

The 5' part of the sequence that was amplified with the primer extension procedure is shown below:

CCACTCCGGGCTTGGGG

The black bold capital letter depicts the nucleotide at which the primer extension stopped. Thus, according to the primer extension assay, the 1st exon of the human GLI3 starts with this nucleotide.

There is a strong prediction for an initiator site found exactly in this sequence, CCACTCC, which includes one nucleotide more (red bold letter) than the primer extension showed (Jacobsen, 1996). This GLI3 start site resides 13 nt downstream to the site where the 5' RACE PCR stopped and 123 nt upstream of the ATG with which the ORF of the GLI3 starts. Interestingly, a transcript with almost the same length (one nt less) was found also with the 5' RACE PCR.

5.3.3 Cloning of promoter constructs

In order to be able to test the complete promoter region in transient cell culture assays for their regulatory activity on a luciferase reporter gene, two promoter fragments were cloned. The promoter elements used in this thesis are depicted in figure 5-7.

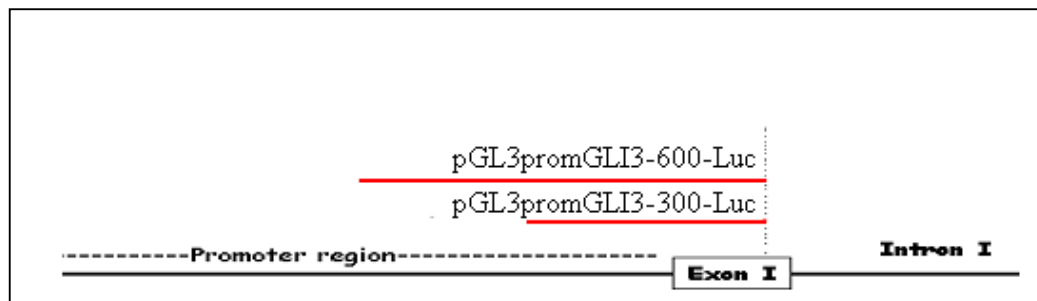


Figure 5-7: Location of GLI3 promoter elements

The GLI3 5' Region including exon I and part of intron I is depicted. The constructs used in this work are shown in red

The corresponding promoter regions including the published 1st exon were amplified with PCR. The forward primer (“Ex1cut for”) anneals 397 bp upstream of the published 1st exon for the first construct and 658 bp upstream of the published 1st exon (primer “Ex1cut forwnew”) for the second construct. The reverse primer (“Ex1cut rev”) starts 4 nt before the end of the 1st exon for both constructs. The primers

included XhoI restriction sites. The fragments used to generate the promoter constructs are shown in figure 5-8.

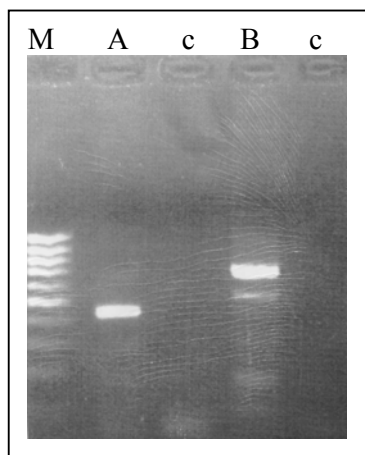


Figure 5-8: Electrophoresis of the promoter PCR products:
PCR products used to generate the 1st (A) and the 2nd (B) construct. c: PCR control; M: 100bp Marker

These PCR products were digested with the XhoI restriction enzyme and were cloned into the respective site of the digested pGL3basic reporter vector. The 1st construct is named “pGL3-promGLI3-300-luc” and the 2nd construct “pGL3-promGLI3-600-luc” (figure 5-9). Isolated clones were sequenced to confirm the presence of the correct insert in the correct orientation.

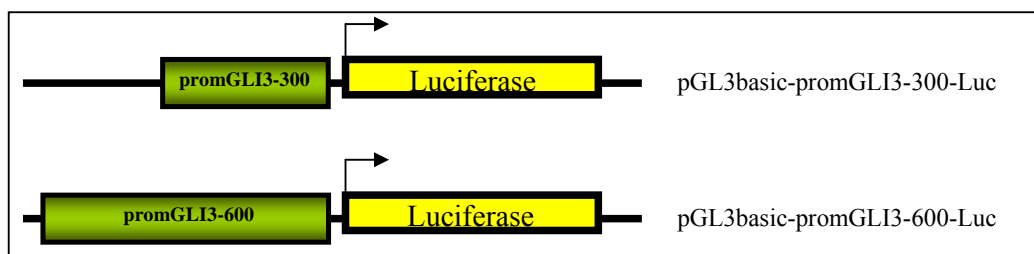


Figure 5-9: Promoter constructs for transient transfection assays:

The two promoter constructs were cloned in front of the luciferase gene of the pGL3basic-luc vector. The arrow depicts the starting ATG of the luciferase reporter gene.

5.3.4 Analysis of the capability of the promoter constructs to drive reporter gene expression in cell culture

pGL3-promGLI3-300-luc and pGL3-promGLI3-600-luc were tested for their ability to drive luciferase reporter expression in cell cultures and compared to the SV40 promoter. The constructs were transiently transfected for two days into H661, a

non-small cellular lung cancer cell line expressing endogenous GLI3. After 48hrs, the cells were lysed and the amount of the produced luciferase from each construct was measured in a dual luciferase assay and expressed in light units (LU). The activities of the constructs are shown in figure 5-10.

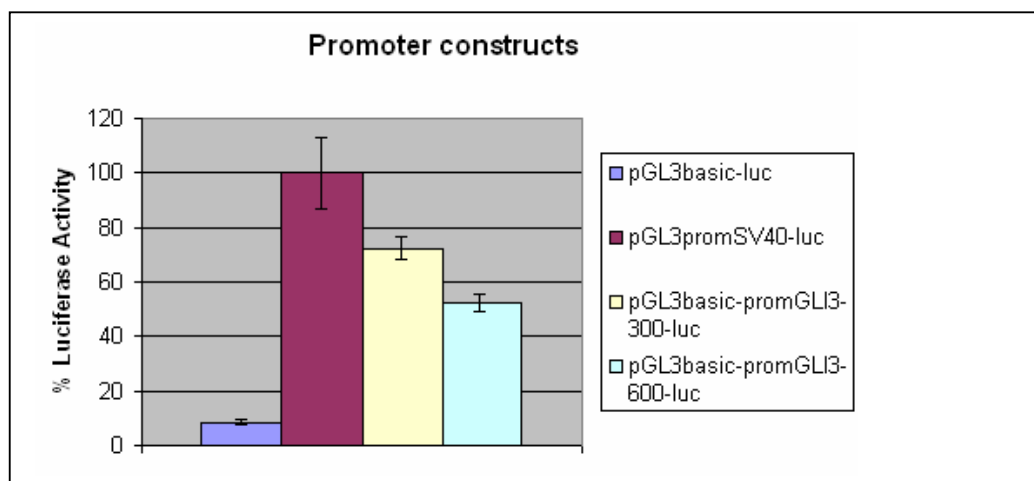


Figure 5-10: Capability of promoter constructs to drive luciferase reporter expression

The pGL3basic construct, which does not contain any element to control its expression, is the vector in which the promoter fragments have been cloned and serves as a negative control. pGL3promSV40-luc is pGL3basic containing in addition the SV40 promoter. Its activity is set to 100%. The clone containing the smaller GLI3 promoter fragment (pGL3basic-promGLI3(300)-luc) shows higher activity than the larger fragment of the GLI3 promoter (pGL3basic-promGLI3(600)-luc).

Since pGL3basic-promGLI3(300)-luc showed a higher activity than the larger fragment, it was used preferentially for the subsequent experiments and was regarded as the *GLI3* minimal promoter that can drive sufficient expression of the reporter gene. To compare its activity on different cellular backgrounds, its ability to drive luciferase expression was tested 48 hrs after transfection also in U2OS cells, which are osteosarcoma cells and express GLI3, and in the H441 cell line which are Clara cells, non-small cellular lung cancer cells and do not show endogenous GLI3 expression. The potential to drive reporter expression is shown in figure 5-11.

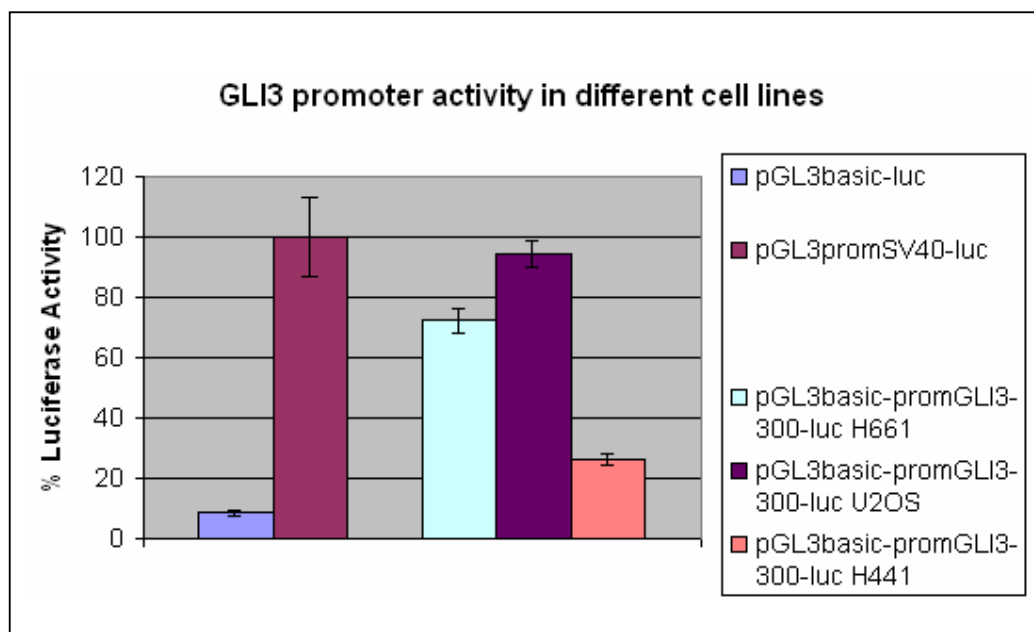


Figure 5-11: Activity of a minimal GLI3 promoter in cultured cells

The activity of the pGL3basic-promGLI3-300-luc construct is slightly higher in the U2OS cells, which are osteosarcoma cells and express GLI3, and decreased in the H441 cell line, which are Clara cells, non-small cellular lung cancer cells and without endogenous GLI3 expression.

5.3.5 Search for critical transcription factor binding sites in the *GLI3* minimal promoter

The sequence of the GLI3 minimal promoter was searched for possible transcription factor (TF) binding sites, which enhance or repress its activity using the TESS online program (<http://www.cbil.upenn.edu/cgi-bin/tess/tess>). TESS is a web tool for predicting transcription factor binding sites in DNA sequences. It can identify binding sites using site or consensus strings and positional weight matrices from the TRANSFAC, IMD, and the CBIL-GibbsMat database. The search was filtered to show only the human TF binding sites (figure 5-12).

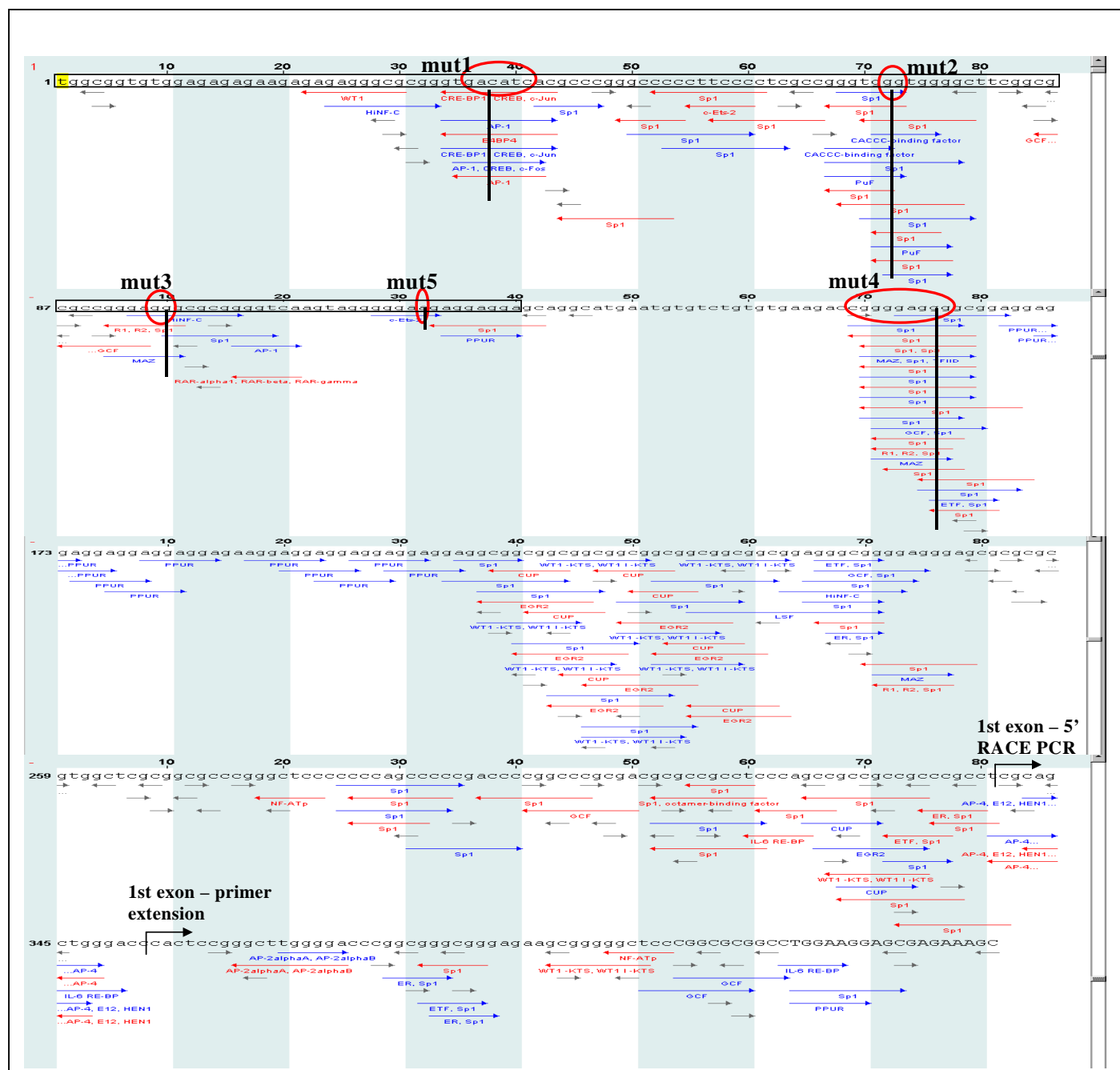


Figure 5-12: Potential TF binding sites in the GLI3 minimal promoter

The possible TF binding sites of the GLI3 minimal promoter are shown with red or blue arrows depending on their orientation. In red circles are the five mutations (mut1 to mut5), which have been introduced in order to test their influence in the activity of the promoter. The black vertical lines show the TF binding sites that could not be predicted anymore in the mutated sequence. The sequence that was removed to create a truncated minimal promoter is shown in the black box. The start of the 1st exon according to the 5' RACE PCR and the Primer Extension analysis is indicated.

Five different sites were mutated using the QuickChange Site-Directed Mutagenesis Kit of Stratagene (mut 1 to 5). The primers (“prommut1F” & “prommut1R”, “prommut2newF” & “prommut2newR”, “prommut3newF” & “prommut3newR”, “prommut4F” & “prommut4R”, “prommut5newF” & “prommut5newR”, for the mutations 1 – 5 respectively) for creating the mutations were designed in such a way that the TF binding sites could not be predicted anymore by the TESS tool and that no other new TF sites were created (prommut 1 to 5). Each mutation was tested individually for its effect on the activity of the pGL3basic-promGLI3-300-luc construct by transiently transfecting it after culture for 48 hrs in the GLI3-expressing H661 cell line.

In addition the GLI3 promoter was truncated at its 5' end by 126 bp in order to define more strictly the region, which is important for its activity. By PCR amplification (primers “Ex1cut rev” & “Ex1promAtrunc F”) the truncated promoter was amplified and cloned in the XhoI site of the pGL3basic-luc vector. This construct was purified and was transiently transfected in H661 cells to test its activity.

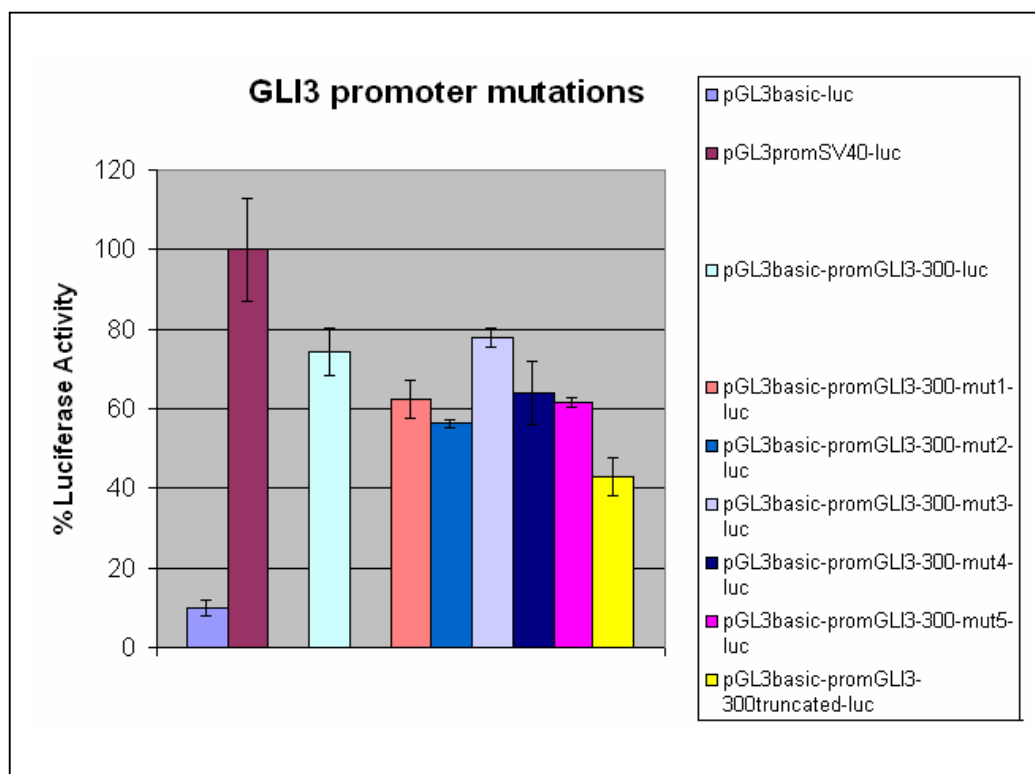


Figure 5-13: Capability of mutated minimal *GLI3* promoters to activate reporter gene expression in H661 cells

The majority of the mutations had only a minor negative effect on the activity of the pGL3basic-prom*GLI3*-300-luc. Mutation 4 did not have any effect at all. The truncated construct showed a weaker activity than the original minimal promoter, but still did not lose its activity.

The site directed mutagenesis of the sites 1, 2, and 5 each had a minor effect on the activity of the *GLI3* minimal promoter. However, the truncation of its upstream sequence reduced its activity by half (figure 5-13). Obviously, this effect is not due to the loss of function of a single site from the tested ones.

Two nucleotides of the initiator site identified with the primer extension method were also altered (primers “InrmutF” & “InrmutR”) to test if this element plays an important role in the promoter activity. The sequence of the initiator site and the adjacent 1st exon according to the primer extension are shown below.

CCACTCC**ggg**ctt**ggg**gaccc**ggc**gggcgggagaagcgggggctccc**ggc**gcggcctggaaggagcgagaa
agcaaaa

The initiator site is shown in capital letters and each of the nucleotides in bold italics was exchanged with a T to create the mutation. The adjacent first exon is shown in small letters. This construct was purified and was transiently transfected in H661 cells to test its activity. However, no difference in the activity could be observed between the mutated (pGL3basic-Inrmut) and the un-mutated construct (pGL3basic-prom*GLI3*-300-luc) (figure 5-14).

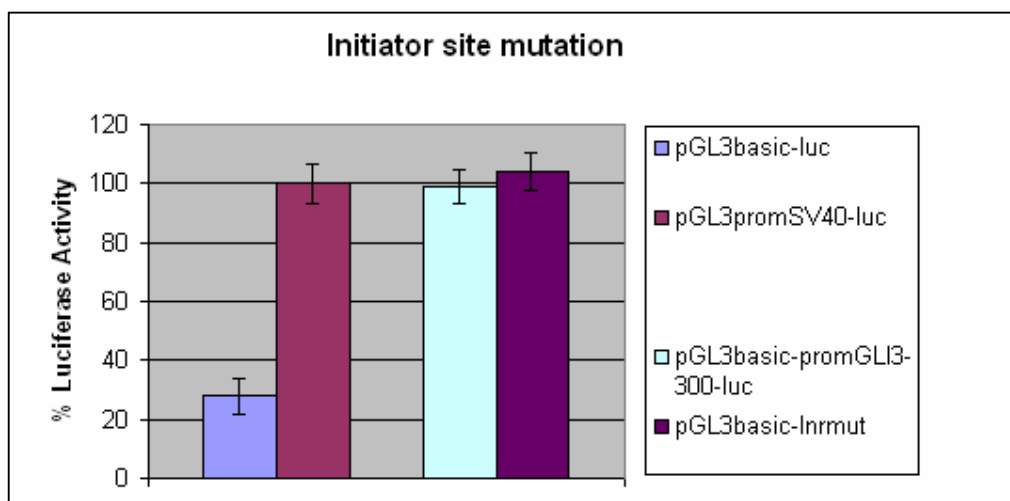


Figure 5-14: Capability of mutated initiator site of *GLI3* promoter to activate reporter gene expression in H661 cells

The pGL3basic-Inrmut construct containing the mutated initiator site predicted by the primer extension does not affect the activity of the *GLI3* minimal promoter.

5.4 Identification of conserved elements downstream of *GLI3* promoter

A 4 kb region downstream of the 1st exon shows high similarity (>75%) between the human and mouse genomic sequence of *GLI3*. The similarities were identified using the VISTA alignment program (<http://genome.lbl.gov/vista/index.shtml>), which calculates in a window of 60-100bp the similarity between given sequences and finally provides the user with a graph in which the conservancy is depicted in figure 5-15.

A 4 kb region with high homology between the species is shown with a black arrow. It resides approximately 1.9 kb downstream of the first exon of human *GLI3*, inside the first intron. The restriction site map was used to select suitable fragments for cloning into a reporter vector and for testing of their regulatory potential.

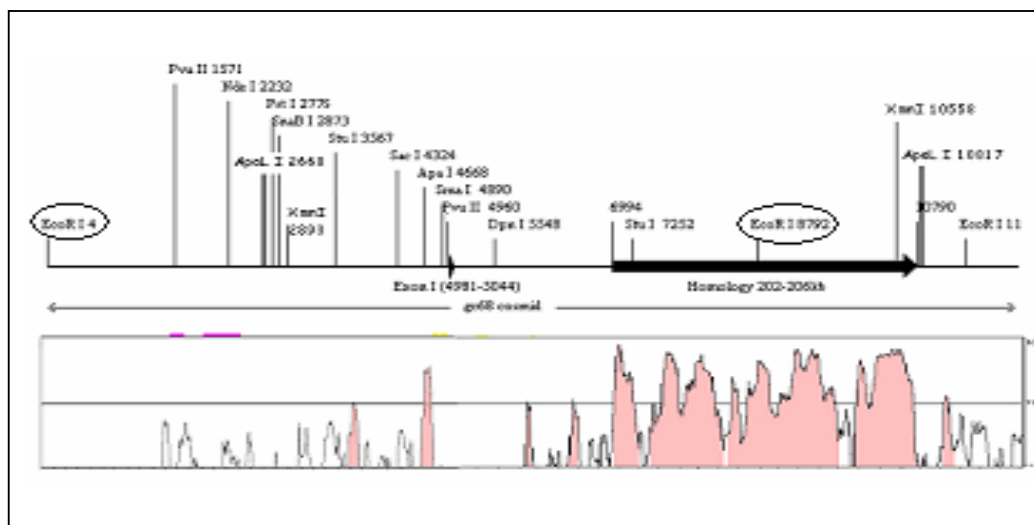


Figure 5-15: Highly conserved genomic sequences around GLI3 exon 1

Above: restriction map of a cosmid encompassing the GLI3 exon 1, the upstream region, and part of intron 1. Below: Vista plot comparing this human sequence with the homologous mouse sequence. Regions of homology higher than 75% are red.

A 4 kb region with high homology between the species is shown with a black arrow. It resides approximately 1.9 kb downstream of the first exon of human GLI3 (small arrow), inside the first intron. The fragment between the circled EcoRI sites was used as a template for cloning.

Such highly conserved non-coding regions may play a regulatory role. Therefore it was decided to clone parts of this region and test their regulatory activity on the luciferase reporter gene.

The circled EcoRI sites in figure 5-15 encompass an 8.8 kb fragment that was used for enzymatic digestions to isolate and clone two different fragments shown in figure 5-16.

The enzymes StuI and DpnI were chosen to digest the 8.8 kb fragment as these enzymes produce blunt ends and allow the inserts to be cloned into the SmaI site of the pGL3basic vector. Blunt end ligation was chosen because no other suitable enzymes could be found for digestion and cloning.

StuI digestion of the 8.8 kb fragment generated a 3.7 kb product and StuI/DpnI digestion resulted in a 1.97 kb element.

The 3.7 kb fragment includes 1.5 kb upstream of exon 1 of human GLI3, the 1st exon, and 2.2 kb downstream into the 1st intron, thus including the potential minimal promoter and the first peak of the high homology region marked in figure 5-16.

The 1.97 kb fragment contains 1.5 kb upstream of 1st exon of human GLI3, the 1st exon and 0.5 kb from the 1st intron.

The two fragments have the same upstream sequence, but one of them lacks the sequence corresponding to the 1st peak of the high homology region.

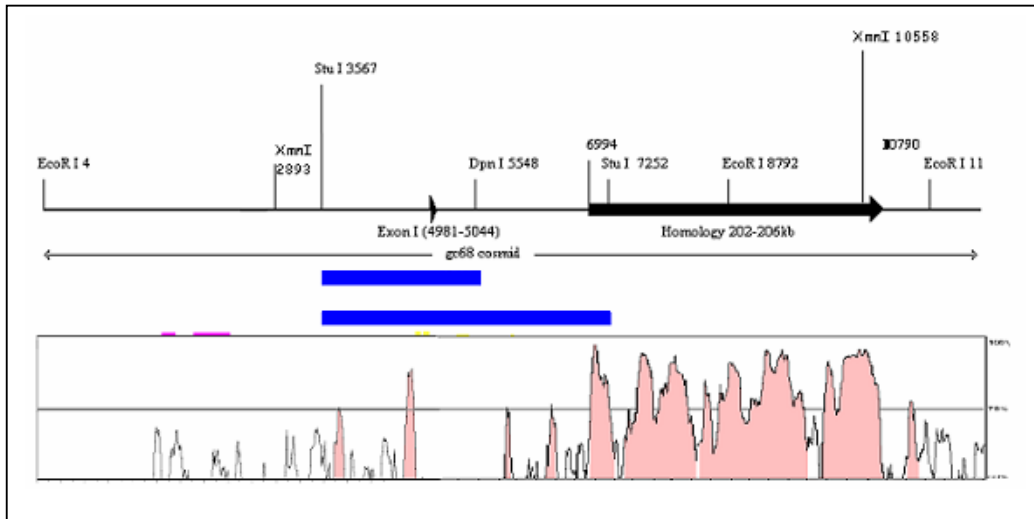


Figure 5-16: Subcloning of sequences around GLI3 exon 1

The blue lines show the regions that were isolated from an 8,8 kb EcoRI fragment and cloned into pGL3basic to be tested for their regulatory potential. Both originate at the same 5' upstream StuI site but have different 3' ends.

The blunt-end digested fragments were isolated and cloned in the SmaI, CIP treated pGL3basic reporter vector. Positive clones with the correct orientation of inserts were identified and tested for their ability to drive expression of the luciferase reporter gene after transfection into the H661 cell line and culturing for 2 days (figure 5-17).

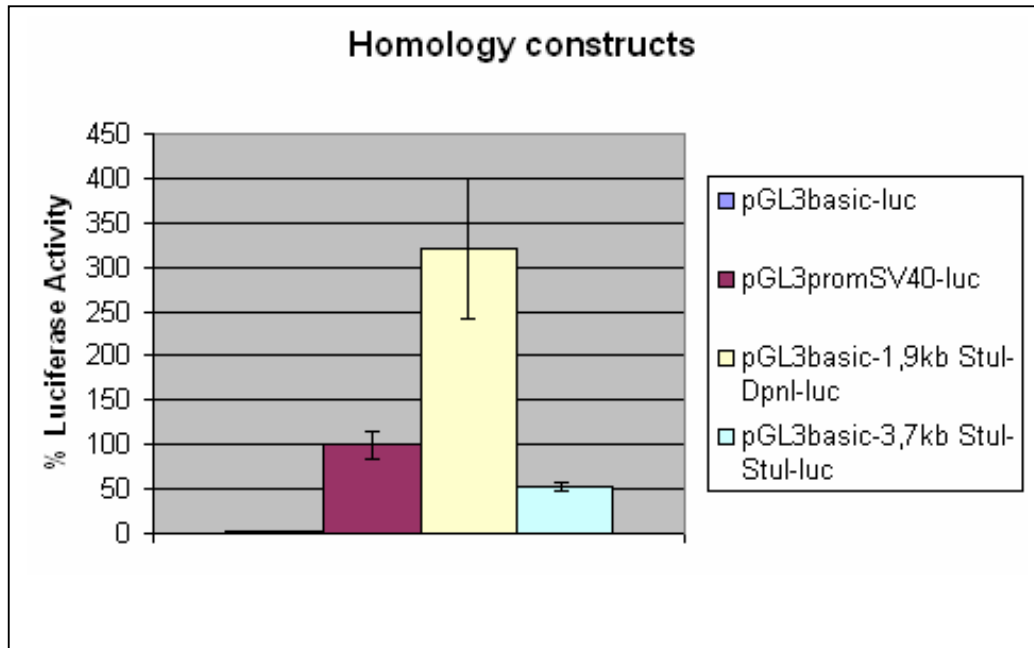


Figure 5-17: Activating potential of GLI3 5' sequences

Luciferase activity in cells transfected with cloned GLI3 5' fragments. The 1.9 kb construct shows a very high potential of driving the expression of the luciferase reporter gene, whereas the 3.7 kb construct, which includes the 1.9 kb region and additional downstream sequence, has a low activity.

The results show a very high activating potential of the short fragment in comparison to the SV40 promoter and in particular in comparison to the larger fragment. Addition of extra sequence from intron 1 down-regulates the activity considerably.

5.5 Highly conserved intronic regions of GLI3

By comparing orthologs between evolutionary distant species, the most conserved regions can be identified. To detect further regions of high conservancy between distantly related species, which might pinpoint regulatory elements, the genomic sequence of human GLI3 was compared with the VISTA alignment program with its ortholog sequence of *Mus musculus* and *Fugu rubripes*. For the human-mouse comparison, 200 kb upstream and downstream of the GLI3 gene were included in the comparison. The fugu sequence was available with 24 kb upstream and 2.5 kb downstream of the gene for the comparison. The fugu GLI3 sequence was kindly provided by Dr. Greg Elgar, London, who also identified fugu sequence elements,

which showed high conservancy when compared with the human and mouse GLI3. The comparison between the different GLI3 orthologs identified numerous highly conserved regions (figure 5-18). It is obvious, that the exonic sequences are highly conserved, however there is also high conservancy between a great number of intronic sites in the human-mouse comparison. The number of non-exonic regions of high homology between all three species is considerably smaller and mainly restricted to intronic areas. Three of those sequence elements were chosen to be tested for their regulatory potential based on their size and their high degree of conservancy in all three species.

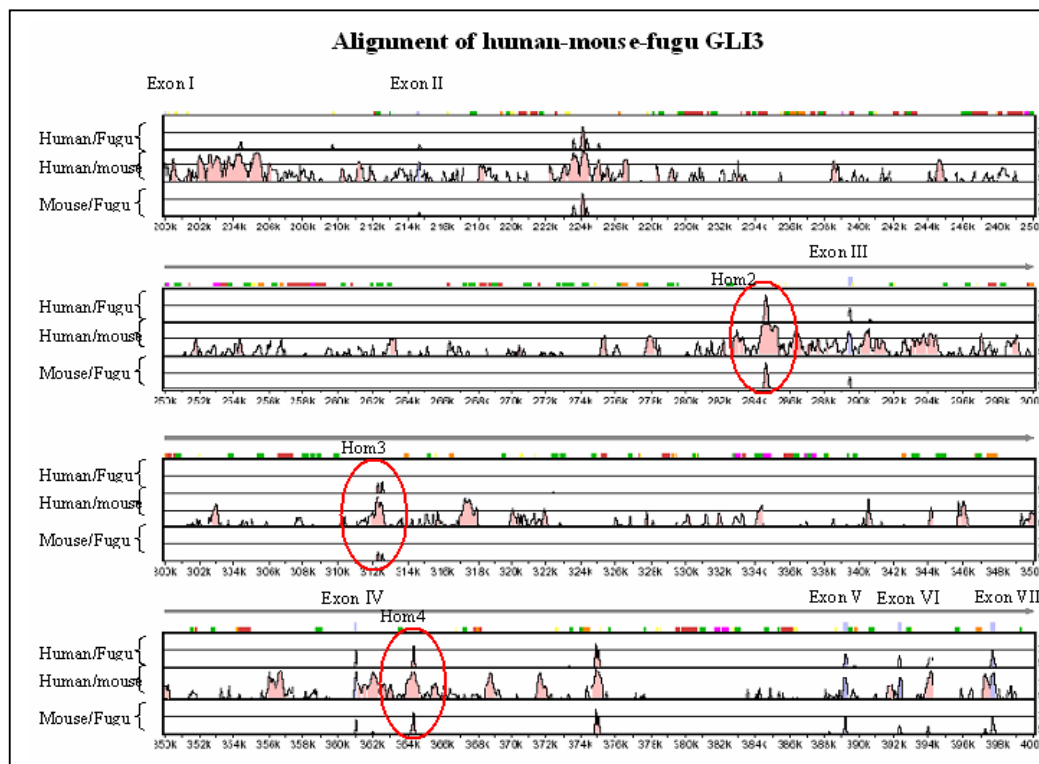


Figure 5-18: Vista plot comparing exon I to intron VII of the GLI3 genes of fugu, mouse, and human

Exons and the peaks representing their sites of homology are shown in light blue. The colored blocks above the plot lines indicate different types of repetitive elements. Sequence homologies of > 50% are indicated as red peaks. Three intronic regions Hom2, Hom3, Hom4, conserved in all three species, were selected for further study (red circled peaks).

The first of these homology regions, Hom2, is a 434bp segment residing approximately 70 kb downstream of exon 2, in the 2nd intron of human GLI3. It shows

100% identity between human and mouse and more than 80% between human and *fugu*.

The second homology region, Hom3, is 386 bp long and begins approximately 23 kb downstream of the 3rd exon, in the 3rd intron of human GLI3. It shows more than 80% identity between human and mouse and around 70% between human and *fugu*.

Finally, the third homology, named Hom4, is a 172 bp long region residing approximately 3 kb downstream of the 4th exon, in the 4th intron of human GLI3. It shows around 90% identity between human and mouse and more than 75% between human and *fugu*.

These three homologous regions were PCR amplified (figure 5-19) from genomic DNA of a healthy person with primers containing HindIII restriction site tags (primers “Hom2New F” & “Hom2New R”, “Hom3fornew” & “Hom3 rev” and “Hom4forwnew” & “Hom4 rev”), and then cloned in the HindIII site of the pGL3basic vector. They were also amplified with the same primers containing KpnI restriction site tags (“Hom2NewKpnI F” & “Hom2NewKpnI R”, “Hom3KpnIfor” & “Hom3KpnIrev” and “Hom4KpnIfor” & “Hom4KpnIrev”) and cloned in the KpnI site of the pGL3basic-promGLI3-300-luc construct (figure 5-20).

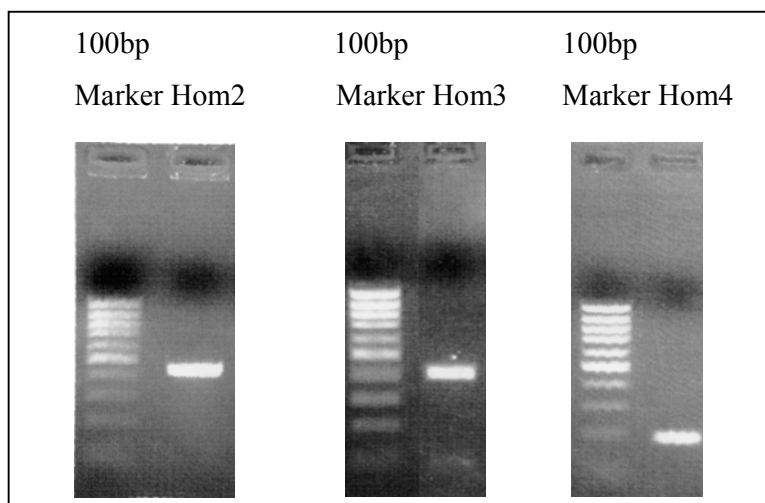


Figure 5-19: PCR products used to clone regions of high conservancy within GLI3

The PCR products of the Hom2, Hom3, and Hom4 are shown after electrophoresis in 2% agarose gels. The size of the bands is slightly larger than the actual length of each homologous region because of the restriction site tags on the primers necessary for the cloning in reporter vectors.

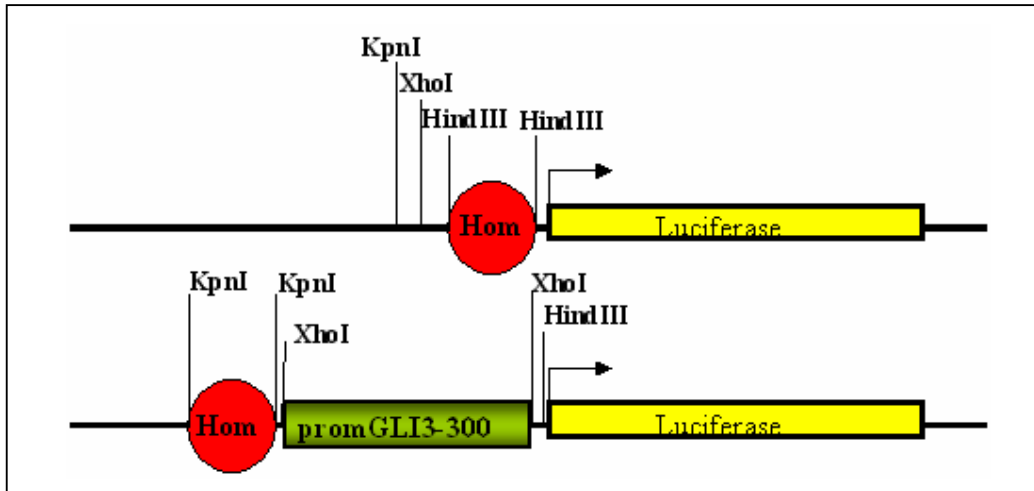


Figure 5-20: Homology constructs

Schematical representation of the homology constructs created. The 3 homologies were cloned in the pGL3basic-luc vector, upstream of the luciferase gene in the HindIII site and in the pGL3basic-promGLI3-300-luc construct in the KpnI site upstream of the minimal *GLI3* promoter in both orientations. The *GLI3* promoter is cloned in the XhoI site and the translation start of the luciferase is depicted with a black arrow.

The constructs were used to test the regulatory activity of the three conserved regions by themselves and in combination with the minimal *GLI3* promoter.

In order to test their ability to drive expression of the luciferase reporter gene, constructs were isolated who had the homologous regions in positive and negative orientation cloned in the pGL3basic and in the pGL3-promGLI3-300-luc vectors. These constructs were sequenced to exclude unwanted mutated fragments and were transiently transfected for two days in the H661 cells, the cell line that expresses endogenous *GLI3* (figure 5-21, 5-22).

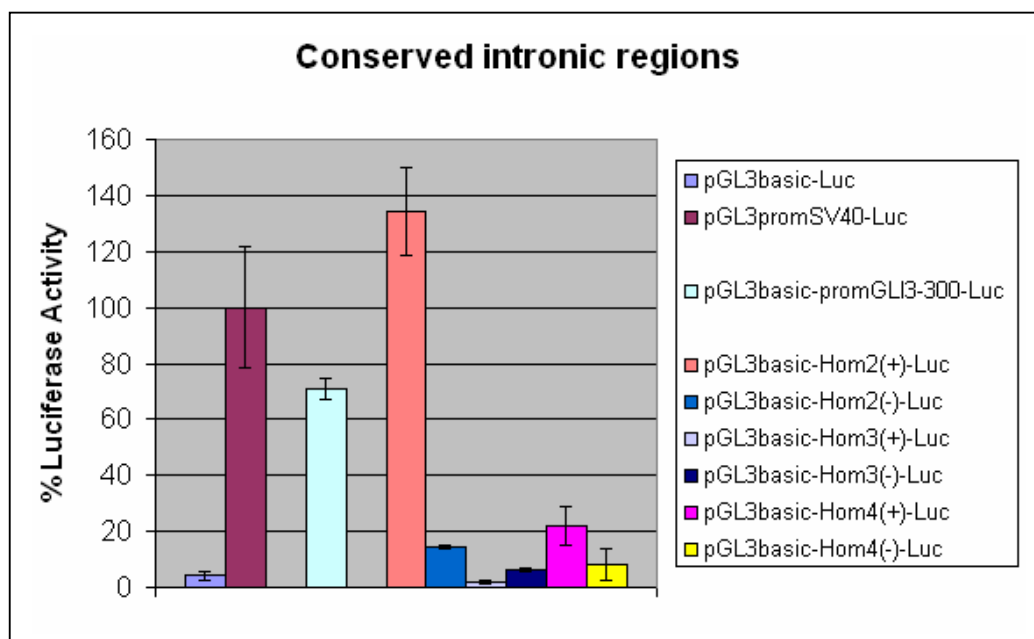


Figure 5-21: Expression of a luciferase reporter under the control of different regulatory elements from the GLI3 region

The luciferase activity of the transfected constructs is compared to the activity of the reporter under the control of the SV40 promoter. Homology 2 by itself shows an activation of the luciferase reporter gene in the positive orientation only. The other homologies by themselves do not seem to have an activity in either orientation.

The homology regions influenced the reporter activity:

Hom2 conserved region showed a very high up regulatory potential when cloned without additional promoter in (+) orientation upstream of the luciferase gene whereas in the (-) orientation it didn't show any activity. In the first 3 times that its activity was measured (3 assays each time), this homology had very high activity as shown in figure 5-21 but after that its activity was fixed down at 40-50% of the SV40 activity.

Hom3 and **Hom4** didn't show any significant activity in either orientation when cloned by themselves in the pGL3basic-luc vector.

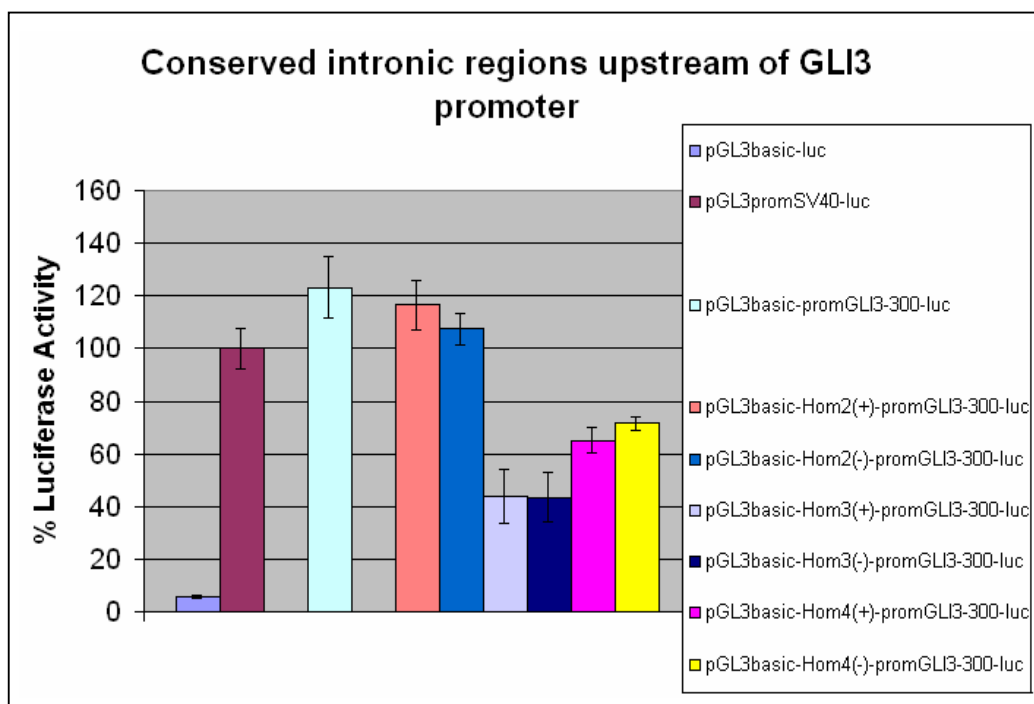


Figure 5-22: Expression of a luciferase reporter under the control of different regulatory elements from the *GLI3* region

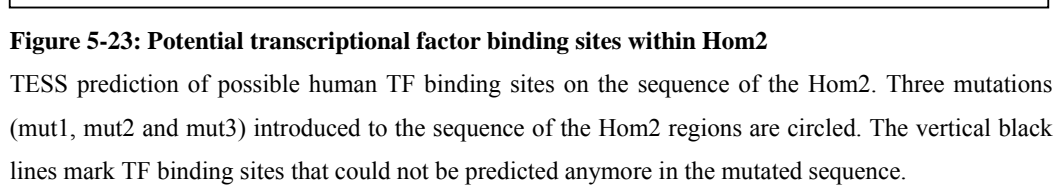
The luciferase activity of the transfected constructs of the 3 homologies cloned upstream of the minimal *GLI3* promoter is compared to the activity of the reporter under the control of the SV40 promoter. Homology 2 in either orientation doesn't show any effect whereas the other two homologies down-regulate the *GLI3* promoter in both orientations, with Hom3 showing stronger negative effect than Hom4.

The homology regions influenced the reporter activity:

Hom2 didn't have any positive or negative effect when cloned in either orientation upstream of the *GLI3* minimal promoter.

Hom3 and **Hom4** in either orientation reduced the activity of the *GLI3* minimal promoter when cloned upstream to it down to 40% and to 65 – 70% of the SV40 activity respectively.

Therefore it was further analyzed for potent transcription factor (TF) binding sites using the TESS program. The search was filtered so that only human binding sites are shown in figure 5-23.



TESS prediction of possible human TF binding sites on the sequence of the Hom2. Three mutations (mut1, mut2 and mut3) introduced to the sequence of the Hom2 regions are circled. The vertical black lines mark TF binding sites that could not be predicted anymore in the mutated sequence.

In the case of mut1, a mutation has been created in the 161-163 nt of Hom2 exchanging aaa with tcc (primers “Hom2mut1F” & “Hom2mut1R”).

In mut2 a mutation has been created in the 180-185nt of Hom2 exchanging aaggaa with ggatcc (primers “Hom2mut2F” & “Hom2mut2R”).

In mut3 a mutation has been created in the 215-219 nt of Hom2 exchanging aggaag with gatcc (primers “Hom2mut3F” & “Hom2mut3R”).

In all cases, the exchange of the target base pairs resulted in the loss of the prediction of the potent TF binding sites by the TESS program and in the creation of a BamHI site for easier detection of the mutated constructs. The novel constructs were also verified by DNA sequencing.

5.5.2 Reporter gene regulation by Hom2 mutants in H661 cells

To test the effect of the mutations, the altered Hom2 constructs were transfected transiently into the GLI3 expressing H661 cell line, and the luciferase activity was measured after 48 h in culture. Each of the three mutations blocked the up-regulatory potential of the Hom2 region (figure 5-24).

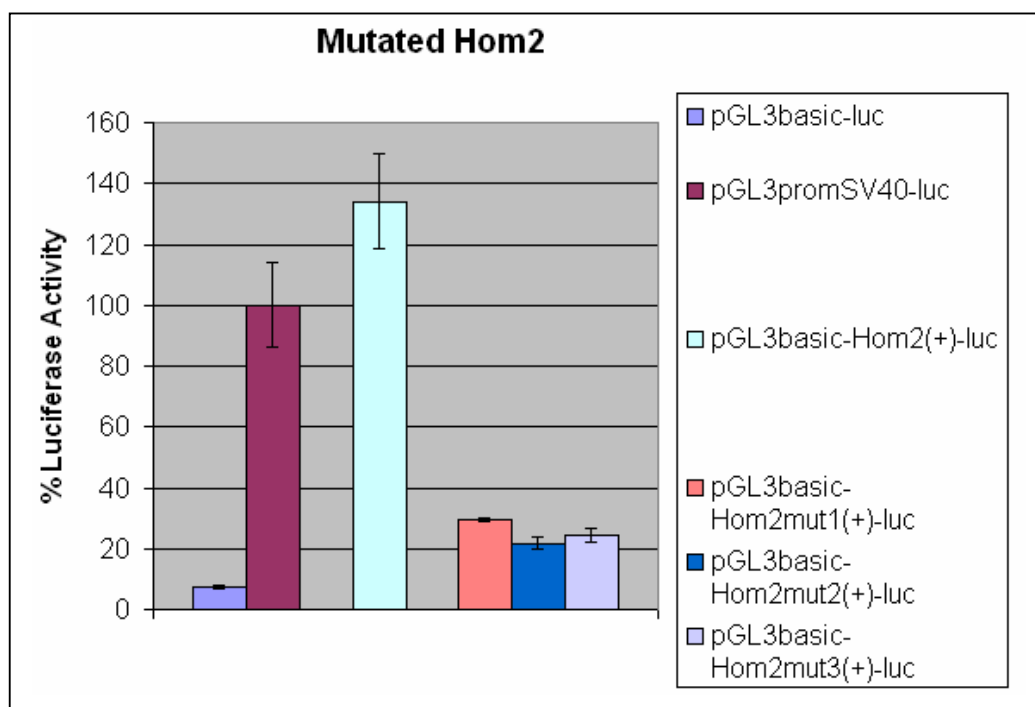


Figure 5-24: Luciferase assay showing the potential of mutated Hom2 elements to activate transcription of a reporter gene

Each of the three mutations decreases the potential of the pGL3basic-Hom2(+)-luc construct to induce luciferase gene transcription in H661 cells considerably.

5.5.3 Reporter gene regulation by Hom2 in different cell lines

The usage of a potential cis regulatory element of transcription in cells depends on the presence of the appropriate factors binding to this segment. To elucidate the role of Hom2 region as a GLI3-specific regulatory element, the pGL3basic-Hom2(+)-luc construct was tested in two other cell lines, U2OS and H441.

The construct was transiently transfected for 48hr in all three cell lines and, after 48 hrs in culture, luciferase was quantified, and so the activity of the same construct in the different cell lines was measured (figure 5-25).

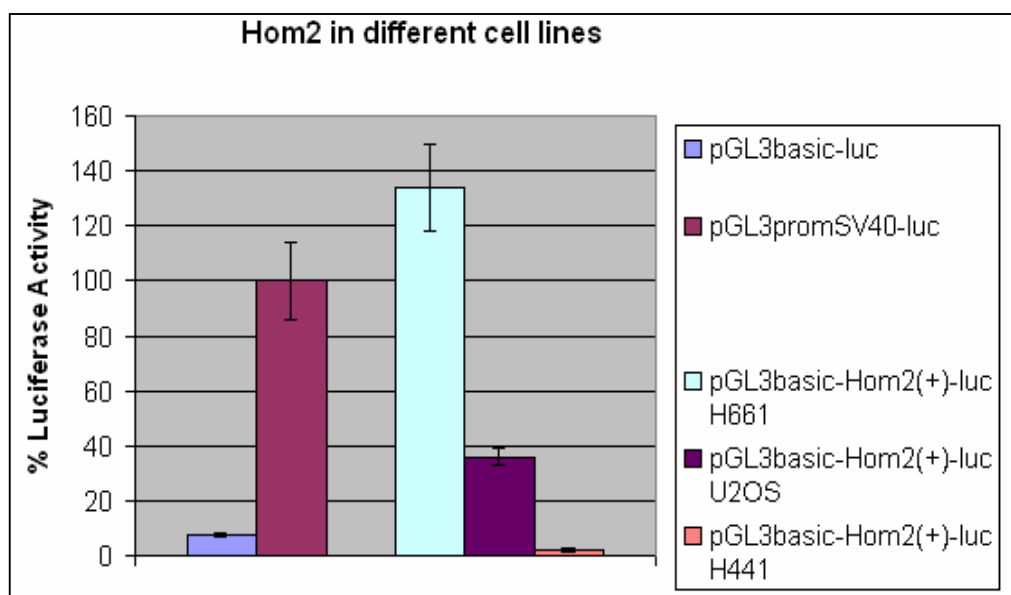


Figure 5-25: Reporter gene regulation by Hom2 mutants in different cell lines

Luciferase activity indicates the ability of the Hom2 region to drive expression of the reporter gene in the construct GL3basic-Hom2(+)-luc in different cell lines. The activity the U2OS cell line is about one quarter of the one observed in H661 whereas no measurable activity is observed in the H441 cell line.

The cell lines that express endogenous GLI3 (H661, U2OS) do not allow the same activity of Hom2. The pGL3basic-Hom2(+)-luc construct shows its highest activity in the H661 line, a relatively low activity in the U2OS cell line and no activity at all in the no GLI3 expressing H441 cell line. Interestingly, the activity of Hom2 region in driving the expression of the reporter gene in the U2OS cell line is similar to the activity of the mutated Hom2 constructs in the H661 cell line.

5.5.4 NFATp as a candidate transcription factor to interact with Hom2

All of the three artificially created mutations in the Hom2 sequence had as a result the decrease of its activity to the same extent. Taking a closer look at the various TF binding sites that are predicted by the TESS program, the only factor that is shared in all three sites is the NFATp factor. To elucidate whether this TF is involved in the regulation of Hom2, first it had to be investigated whether it is expressed in the H661, U2OS and H441 cell lines, in which the Hom2 showed different potential of transcription activity. For this reason, NFATp specific primers were designed to anneal on the cDNA sequence of this gene (“NFATp F” & “NFATp R”). These primers were used for PCR amplification on cDNA that was reverse transcribed from total RNA that was isolated from the three different cell lines (figure 5-26).

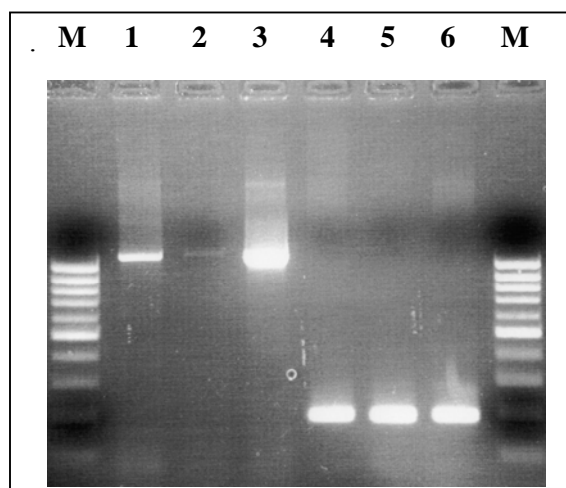


Figure 5-26: RT-PCR of transcription factor NFATp mRNA from different cell lines

1-3: NFATp specific fragments, **4-6:** beta-actin specific fragments; H661 cell line, 1+4; U2OS cell line, 2+5; H441 cell line, 3+6. **M** = 100bp marker

NFATp cDNA could be amplified most strongly in the H441 cell line, in which the Hom2 did not show any activity on reporter gene expression. U2OS, which showed a weak expression of Hom2, allowed a very weak amplification of the NFATp fragment, whereas the same fragment was amplified in a sizeable amount in the H661 cell line, in which the Hom2 fragment showed the highest activity. A fragment of beta-actin cDNA was PCR amplified (“b-actin F” & “b-actin R”) as a control for the efficiency of the reverse transcription of total RNA from all the cell lines and the equal amount of template applied.

The NFATp fragments were sequenced to verify that the amplification was indeed NFATp specific.

5.5.5 Search of a transcriptional start by primer extension analysis on Hom2

Hom2 by itself shows high potential of driving expression of the luciferase gene, and this only in the positive orientation. This independency of any promoter means that the sequence must contain sites that enable the initiation of transcription. Therefore, a primer was designed with a 5' IRD-700 modification annealing within the luciferase gene 71bp downstream of the 3' HindIII site in which Hom2 is cloned. This primer directs towards the 3' end (“Lucexten”) of the Hom2 and was used to synthesize cDNA by RT-PCR from H661 total RNA, which before had been transfected transiently for 48hr with the pGL3basic-Hom2(+)-luc construct. This product was used to identify the transcription start site that is used for the expression of the luciferase from Hom2.

The resulting cDNA was precipitated and run in a sequencing gel next to the sequencing reaction of the pGL3basic-Hom2(+)-luc construct, which was sequenced with the same primer that was used for the primer extension assay (figure 5-27).



Figure 5-27: Primer extension analysis of the Hom2 transcription start site

The primer extension product next to the sequence of the pGL3basic-Hom2(+)-luc construct. The signal from the primer extension band marks the nucleotides with which the cDNA, thus the mRNA, of the reporter construct begins.

Part of the sequence that was amplified with the primer extension procedure is shown below.

GGGTGAATTGAGCTGAGCGGCTAGCTGG

The primer extension resulted in three different bands that correspond to the sequence depicted in bold letters. Since the sequence in the critical region was difficult to read it does not allow the exact identification of the transcription start. However, the data show that there exists a transcription start region within a 13 nt sequence that resides 35 nt upstream of the 3' HindIII cloning site, inside the sequence of Hom2 region (figure 5-28).

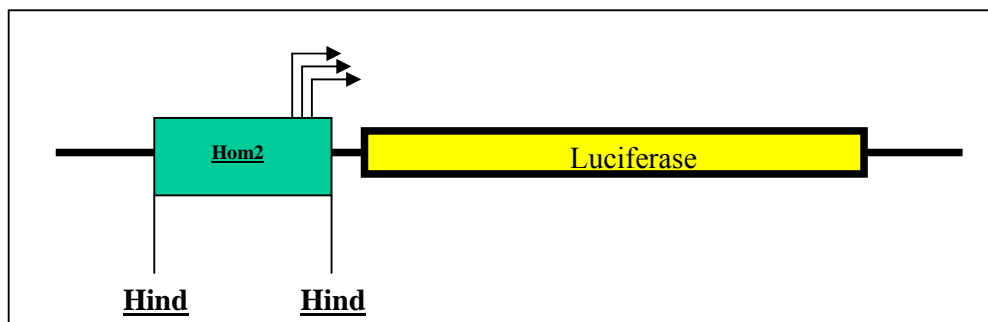


Figure 5-28: pGL3basic-Hom2(+)-luc construct with predicted inr sites

The Hom2 region is cloned in the HindIII sites of the pGL3basic-luc vector, upstream of the luciferase gene. The three black arrows represent the three different sites at the 3' end of Hom2 sequence with which the transcription initiates from this construct according to the primer extension assay.

5.5.6 Deletion mapping of functionally important segments of Hom3

In order to delimit the region that is responsible for the action of Hom3, the fragment, which showed stronger down-regulation of the *GLI3* minimal promoter, its sequence was tested after deletion of either the 5' - or the 3' end. Primers were designed to amplify the Hom3 sequence from its 1st until the 268th nucleotide (5'-Hom3) (primer "3'cutHom3rev") and from its 114th to its last, the 386th nucleotide (primer "5'cutHom3for") (figure 5-29). In this way, two partially overlapping fragments have been amplified, purified and cloned in both orientations in the KpnI site of the pGL3basic-promGLI3-300-luc construct, upstream of the minimal promoter of the human GLI3 gene.

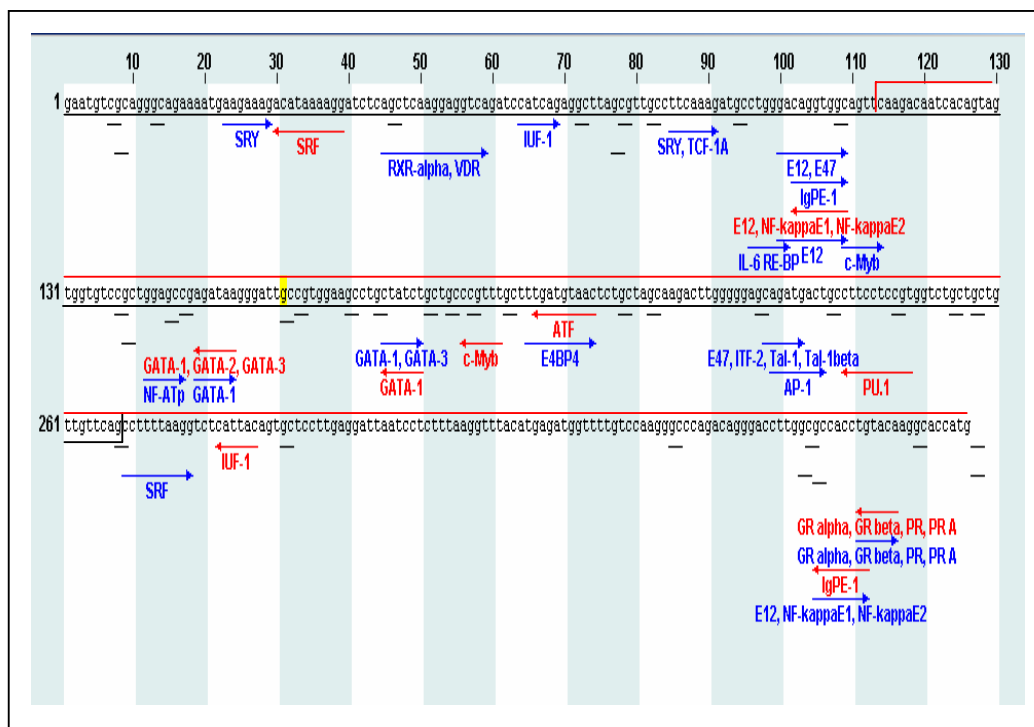


Figure 5-29: The sequence of Hom3, potential TF binding sites, and description of subclones

Prediction of potent TF binding sites in the sequence of the Hom3 region as determined with the TESS program. The sequence was split in two parts overlapping in the boxed area from nt 114 to nt 268.

The desired clones were isolated and transiently transfected and expressed for 48 hrs in the GLI3 expressing H661 cell line (figure 5-30). They delineate three segments, the 5'- and 3' parts present in one clone each, and the middle part present in both constructs.

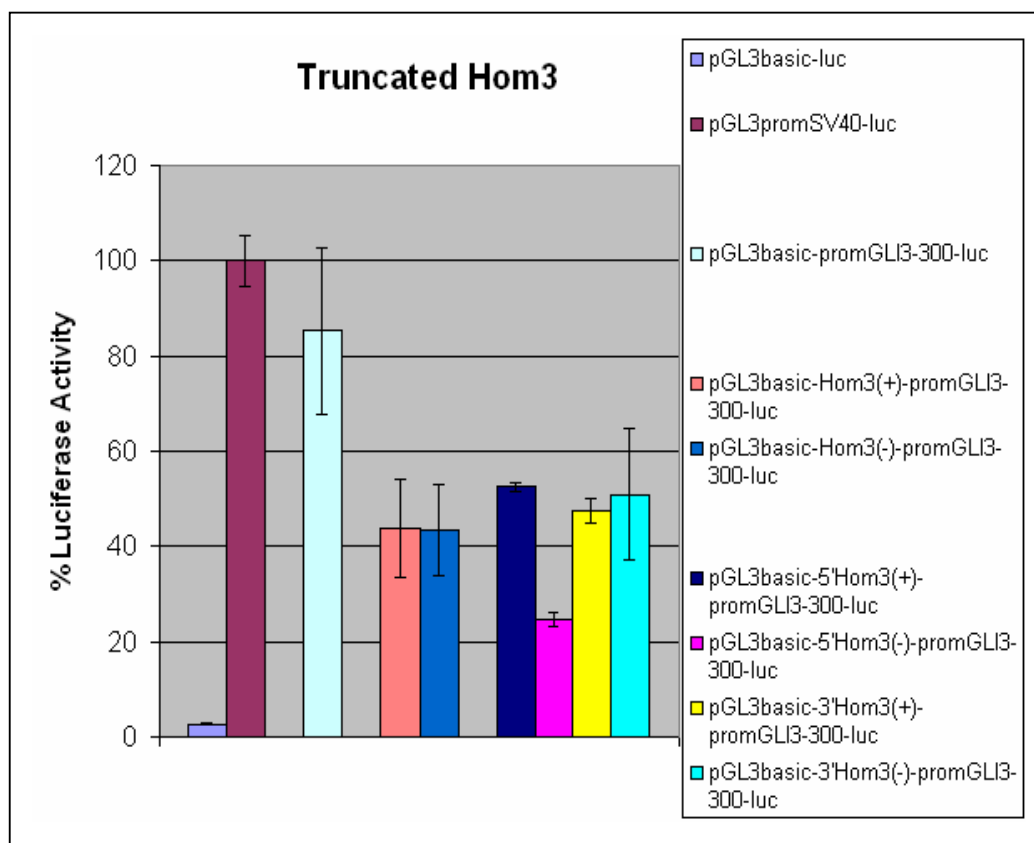


Figure 5-30: Effect of Hom3 truncation

Reporter constructs with a 5'Hom3 fragment and a 3'Hom3 fragment upstream of the minimal GLI3 promoter in a luciferase assay after transient transfection into H661 cells.

Both the 5' and 3' Hom3 constructs have almost the same inhibiting influence on the minimal GLI3 promoter as the whole Hom3 region. They decrease by half the activity of the minimal GLI3 promoter to drive expression of the luciferase reporter gene. Only the pGL3basic-5'Hom3(-)-promGLI3-300-luc construct represses the human GLI3 promoter more than the original Hom3 sequence.

5.6 The potential of the homology regions Hom2, Hom3, and Hom4 to drive reporter gene expression in transgenic mice

Homologies 2, 3 and 4 were cloned in both orientations into the HindIII site of the p1230 vector. This vector carries a lacZ reporter gene under the control of the murine beta-globin promoter, which is inactive in transgenic mice if transfected without additional cis-regulatory sequences. By cloning DNA fragments that are

suspected of containing enhancer elements upstream of the b-globin promoter, one can test if they can regulate positively this promoter and drive expression of the lacZ gene in the tissues for which the enhancer element is specific. The purified constructs were digested with NotI/XhoI so that the DNA fragments containing the homology attached to the b-globin promoter-lacZgene-SV40polyA tail were excised from the p1230 vector.

The positive orientated Hom2 – lacZ construct as well as Hom3 and Hom4 constructs in both orientations were isolated, precipitated, and then injected into the fertilized eggs.

With support by Dr. Esther deGraaff, Rotterdam, the reporter constructs were injected into the male pronuclei of murine fertilized eggs, each. These were then transferred into the oviduct of 3 pseudo-pregnant female mice for Hom2 and of 2 pseudo-pregnant female mice for each orientation of Hom3 and Hom4. In detail, the donor eggs originated from FVB females (approx 3 weeks old and superovulated with hormones), which were crossed with FVB males. The eggs were transferred in the oviduct of CBA females (crossed with sterile CBA males). The numbers of injected eggs for Hom2 were 175 (140 transfers), for Hom3 were 264 (220 transfers) and for Hom4 were 128 (120 transfers).

In this way, the constructs got randomly integrated in the chromosomes of the embryos. Around 20 embryos of the mice carrying the Hom3 and Hom4 construct in each orientation were removed after 11 days. For Hom2, 12 embryos of one mouse were removed after 9 days and of 20 embryos of the other two mice after 11 days. The embryos were stained with 4% paraformaldehyde and stained with X-gal according to the protocol found in the methods section.

Only three 11-days old embryos and one 9-days old embryo that had received the Hom2 construct expressed the reporter as shown by lacZ staining. No staining was observed in transgenic mice with Hom3 and Hom4 constructs. DNA was extracted out of the yolk sac of each transgenic embryo-mouse and was used as a template for PCR, which specifically amplifies the lacZ sequence, to test if the injected transgene was integrated into the host genome (primers “LacZ F” & “LacZ R”) (figure 5-31).

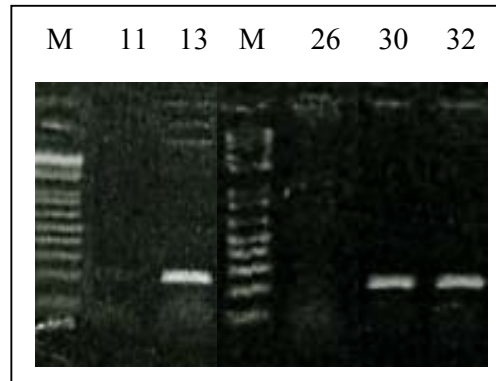


Figure 5-31: PCR gene integration analysis

PCR products of the yolk sac of mouse embryos 11, 13, 26, 30 and 32 obtained with lacZ specific primers. M: 1kb ladder

Embryo 26 does not show a band (and does not show lacZ staining). Embryo 11 has a faint PCR band but shows strong lacZ staining. Mice 13, 30 and 32 also showed lac-Z staining. Out of the 20 11-day old embryos that came out of the Hom2 mice, only 15 were found to contain the lacZ insert in their genome, while out of the 12 nine-days old embryos that came out of the 3rd Hom2 mouse, 9 were identified as positive for carrying the lacZ sequence. That means that only 10-15% of the embryos that had integrated the transgene in their DNA were lacZ stained. Around 75% of the embryos obtained out of the Hom3 and Hom4 injected mice were found to contain the lacZ fragment but none of them showed any staining.

The staining that was observed in mouse embryo 13 (9-days old) as well as in embryos 11, 30 and 32 (11-days old) is shown below. Mice 30 and 32 came from a different mother mouse than mouse 11.

5.6.1 Stained embryos

5.6.1.1 Embryo 13 - 9 days

The only 9-day old embryo that showed lacZ staining is demonstrated in figure 5-32. Strong staining can be observed in the future forebrain region. This region is not yet closed in this developing stage.

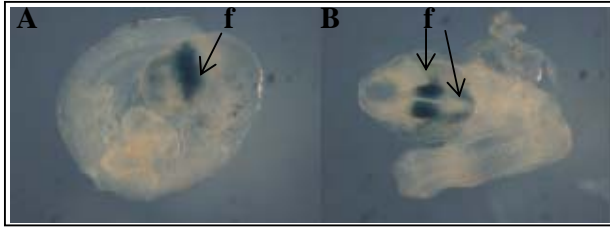


Figure 5-32: LacZ staining of embryo 13 (9-days):

The arrows show the staining in the developing forebrain (**f**) from a side view (**A**) and from a dorsal view (**B**)

5.6.1.2 Embryo 11 – 11 days

The 1st from the three Hom2-lacZ stained 11-days old embryo showed a very intense lacZ staining in the telencephalon and diencephalon, the mandibles, nostrils, and the somites. In addition, a weaker staining was observed in the midbrain, hindbrain and the heart (figure 5-33). This embryo was the only one to show somite staining.

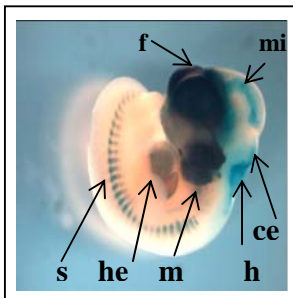


Figure 5-33: Lac-Z staining of embryo 11 (11-days), side view

The staining in the forebrain (**f**), midbrain (**mi**) and hindbrain (**h**), mandibles (**m**), heart (**he**) and somites (**s**) as can be seen in the whole mount embryo preparation. The nostril staining cannot be seen here. From this view it is clearly visible that the staining is most prominent in the telencephalon and diencephalon (forebrain), the mandibles and the somites. The region covering the cerebellar plate (**ce**) of the metencephalon is also stained.

As described before, the staining was very intense in the forebrain region. The region of the telencephalic vesicle and the diencephalon was completely dark, in contrast to the region of the midbrain, which was not stained in such a degree. The region covering the hindbrain and the cerebellar plate of the metencephalon as well as inside

the hinbrain was also stained but not as strong as the forebrain region (previous figure 5-33). In addition, the somites along both sides of the 11-day old embryo and the walls of the heart were stained eminently (figure 5-34).

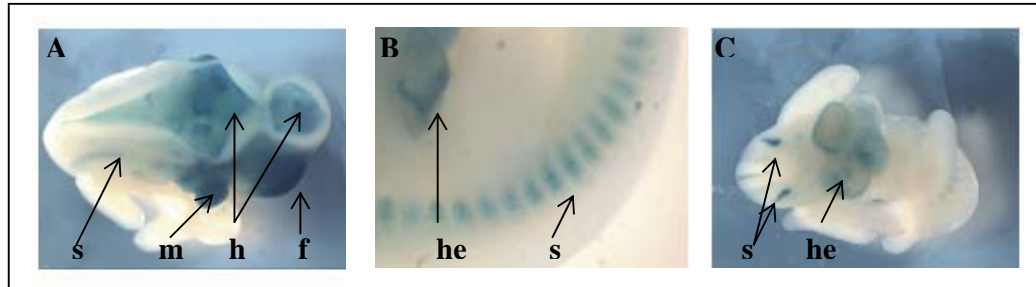


Figure 5-34: Embryo 11, hindbrain and somite staining

In **A** the head has been chopped so that the staining in the hindbrain (**h**) can become visible. The staining of the mandibles (**m**) and the forebrain (**f**) can also be seen. The right somites (**s**) are stained throughout the side of the embryo. In **B** the somites can be seen clearly in the right side of the embryo, regions of the heart are stained as well (**he**). **C** shows the embryo from the top, lacking its head. The somites can be seen in both sides as well as the various parts of the heart.

Regarding the front part of the head, staining can be observed in the olfactory pit and in the medial nasal process while no staining exists in the lateral nasal process. In the mandible, both the maxillary and the mandibular process are stained. In addition, lacZ is expressed also in the developing primary palate (figure 5-35).

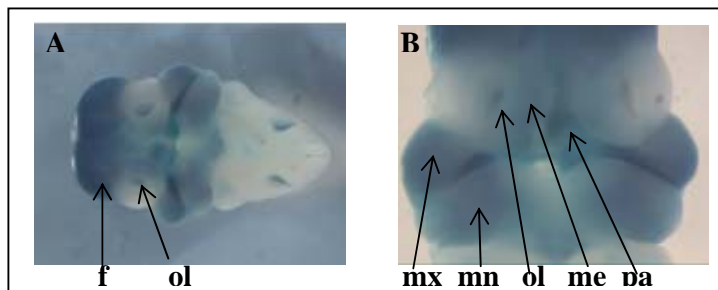


Figure 5-35: Embryo 11, forehead staining

A: The forebrain (**f**), the olfactory pit (**ol**) and the mandibles are shown with lacZ staining. **B:** A magnified look into the same region shows staining in the maxillary (**mx**) and mandibular (**mn**) process, the olfactory pit, the medial nasal process (**me**) and the developing primary palate (**pa**).

Transverse sections from the head of the embryo 11 were analysed for detailed study of the regions and the cells that were stained. The walls of the telencephalon showed an intense staining, the walls of the diencephalon a weaker one (figure 6-9A). At a

more ventral position in the head, staining could be observed in the trigeminal (V) neural crest cells migrating into the maxillary component of the first branchial arch, the maxillary component of the first branchial arch, the mandibular component of first branchial arch, the trigeminal (V) ganglion, and some staining was seen in parts of the wall of the hindbrain (figure 6-9B).

5.6.1.3 Embryo 30 - 11 days

The 2nd from the three Hom2-lacZ stained 11-days old embryos showed again a very intense lacZ staining in the telencephalon and diencephalon and a weaker staining in the midbrain, hindbrain, the mandibles and the nostrils. This embryo did not show somite or heart staining (figure 5-36).

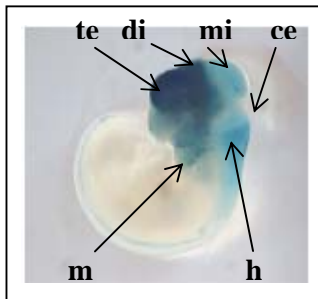


Figure 5-36: Lac-Z staining of embryo 30 (11 days), side view

Lac-Z staining is evident in the diencephalon (**di**) and telencephalon (**te**). It can also be observed in the midbrain (**mi**), the hindbrain (**h**), the cerebellar plate (**ce**) and the mandibles (**m**). The blue stain across the dorsal part of the embryos body most probably is not a product of lacZ expression at this position but a product of dye leakage.

The blue stain observed along the dorsal part of the embryos body is considered to be unspecific, most probably due to leakage of the dye from destroyed brain cells that get carried away in the neural tube. In addition to the forebrain and midbrain, expression of lacZ was observed in the developing hindbrain. By chopping the head at this region it could be seen that the staining is not restricted only to the surface but is present in the interior of the hindbrain (figure 5-37).

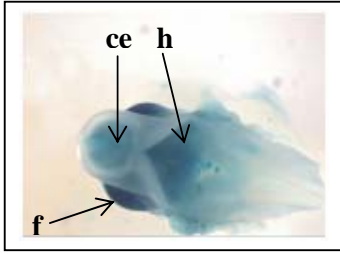


Figure 5-37: Embryo 30 head, dorsal aspect

LacZ staining can be observed in the cerebellar plate and in the hindbrain (**h**) itself. Part of the forebrain (**f**) is also visible.

In the forehead region, the staining is similar to mouse 11. Staining can be observed in the maxillary and mandibular process, in the medial nasal process, the olfactory pit and the developing primary palate. Again, no staining in the lateral nasal process as in mouse 11 (figure 5-38).

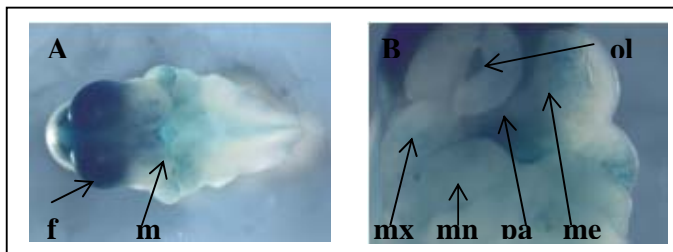


Figure 5-38: Embryo 30, head, ventral aspect

A: The forebrain (**f**) and the mandibles (**m**) are shown with lacZ staining. **B:** A magnified look in the same region shows staining within the maxillary (**mx**) and mandibular (**mn**) process, the olfactory pit (**ol**), the medial nasal process (**me**) and the developing primary palate (**pa**).

5.6.1.4 Embryo 32 - 11 days

The 3rd Hom2-lacZ stained 11-days old embryo showed again a very intense lacZ staining in the telencephalon and diencephalon and a weaker staining in the midbrain, hindbrain, the mandibles and the nostrils (figure 5-39). In addition it showed also staining in the heart and weak staining in the somites, similar to embryo 11, but also in the Wollfian duct and the second brachial arch that develops to cartilage derivatives.

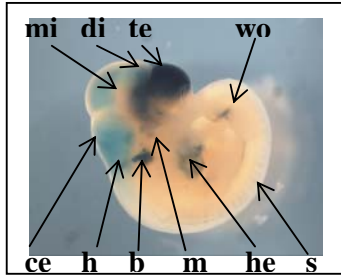


Figure 5-39: Lac-Z staining of embryo 32, (11 days), side view

Lac Z staining can be seen in the diencephalon (**di**) and telencephalon (**te**). It can also be observed in the midbrain (**mi**), the hindbrain (**h**), the cerebellar plate (**ce**), the mandibles (**m**), the second brachial arch (**b**) and the Wolffian duct (**wo**). In addition, lacZ staining appears again in the heart (**he**) and weaker in the somites (**s**).

In the forehead region, the staining is similar to the one observed in the other embryos. The maxillary and mandibular process, the medial nasal process, the olfactory pit and the developing primary palate were lacZ stained. No staining can be observed in the lateral nasal (figure 5-40A). The inner parts of the hindbrain as well as the cerebellar plate were also stained weakly like in the other mice. The second branchial arch was labelled at both sides behind the hindbrain (figure 5-40B)

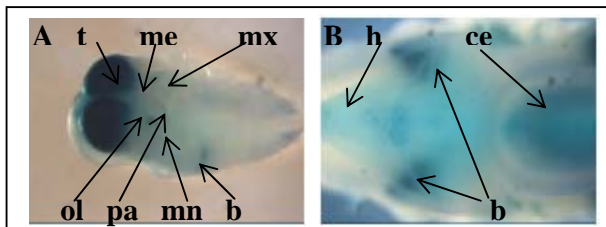


Figure 5-40: Embryo 32, forehead and hindbrain, ventral aspect

A: Ventral view of the forehead of the embryo. Staining can be observed in the telencephalon (**t**), the maxillary (**mx**) and mandibular (**mn**) process, the olfactory pit (**ol**), the medial nasal process (**me**) and the developing primary palate (**pa**). Also part of the right second brachial arch (**b**) is shown with staining. **B:** Superior view of the hindbrain (left part of figure) and the cerebellar plate (**ce**) of the chopped head (right part of the figure). The stained second brachial arch is also visible at both sides behind the brain.

The Wolffian duct is a paired organ found in mammals including humans during embryogenesis. It connects the primitive kidney Wolffian body to the cloaca and serves as the anlage for certain male reproductive organs. This organ was found

stained in the embryo 32. In addition, the heart was stained again as in mouse 11 (figure 5-41).

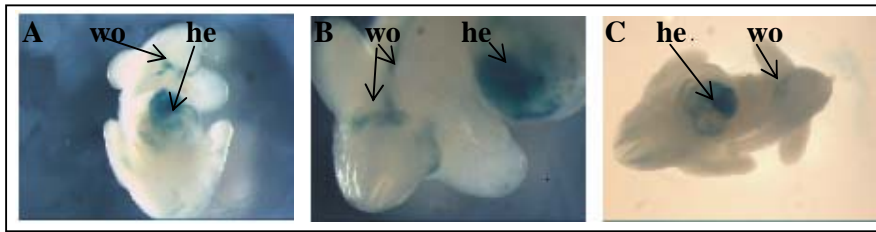


Figure 5-41: Embryo 32, woffian duct and heart staining

A: Ventral view of embryo. Staining can be observed in the heart (**he**) and the Wollfian duct (**wo**). The head and tail were chopped for better visualization of the stained organs. **B:** magnified dorsal view showing part of the heart and the Wollfian duct stained. **C:** Different ventral view of the embryo showing the stained heart and part of the Wollfian duct.

Embryo 32 was the only one to show a faint staining in the upper limbs. The staining was restricted only to the anterior part of the developing limb bud (figure 5-42). Interestingly, this is the region of the limb bud to which the expression of GLI3 is restricted.

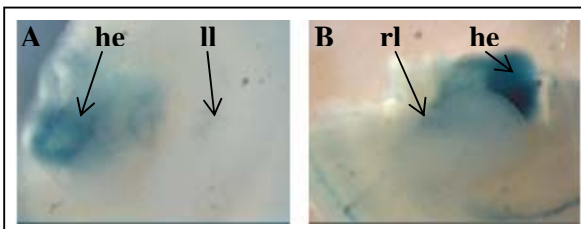


Figure 5-42: Embryo 32, upper limbs staining.

A: Weak lacZ staining in the anterior part of the left developing limb bud (**ll**) left side view. The stained heart can be seen at the left part of this image (**he**). **B:** Weak lacZ staining in the anterior part of the right developing limb bud (**rl**). The stained heart can be seen at the upper part of this figure, where the embryo is lying on its back, the head towards the right side of the image (not shown), right side view.

The comparison of the stained regions in the three 11-days old embryos demonstrates, that Hom2 is an element that can drive specific reporter expression in the fore-, mid- and hindbrain, the somites, the mandibles, parts of the nose and the heart.

6 DISCUSSION

Since the sequencing of the human genome was accomplished by the Human Genome Project, the study of gene regulation and function remains a major task. Elucidating the mechanisms of transcriptional regulation provides insight into the selective expression of genes in different tissues, the differentiation of the cells, and finally the development of complex organisms, such as humans, from a single cell. Since humans are not amenable to experimental genetic variation, mutations leading to altered phenotypes provide valuable hints at normal gene function. They can affect coding sequences, thus impairing normal protein function, or they might disturb promoters and other cis-regulatory elements which control time, location, and quantity of gene expression (Lalioi et al. 1997). Whereas mutations in coding sequences have been studied in detail in a multitude of genes, cis-regulatory sequences have been identified only for few genes in sufficient detail to include them in mutation analysis. The availability of more or less complete genome sequences from many eucaryotes or procaryotes in public databases and of biocomputing tools to compare and annotate these treasures is offering the option to focus the search for regulatory sequences on relatively short, highly conserved sequence intervals outside the coding parts.

To apply this strategy for the analysis of GLI3, a key transcription factor transmitting hedgehog signaling during human embryonal development, the promoter region immediately 5' to the transcriptional start was identified and examined, and other potentially cis-acting sequence elements were studied among highly conserved non-coding elements within GLI3 exons.

Mutation of such elements in patients with GLI3 morphopathies, developmental defects attributed to defective GLI3 function, would give credit to their expected function. It was one task of this thesis to contribute new cases to a resource of probes from patients with appropriate phenotypes, which are not mutated within GLI3 coding sequences.

Based on studies in animals, the paralogous gene GLI2 is expressed near GLI3 in a highly coordinated fashion and appears to share with this factor overlapping

functions. Gli2 and Gli3 gene knockout experiments in the mouse indicate that they may have different roles, but also possess functional redundancy (Mo et al. 1997).

If these results would apply to humans as well, one could expect to observe homologous, even paralogous regulatory sequences in addition to conserved protein sequences. Thus, mutations affecting regulatory or coding sequences of *GLI3* might only result in a recognizable phenotype if GLI2 could not compensate for the missing function of this factor. However, GLI3 haploinsufficiency in humans, in contrast to the mouse, shows a phenotype comparable to the homozygous situation in the animals. Thus, regulation and/or function of human GLI2 might differ from the one encountered in GLI3, a notion supported by homology searches in sequence databases that predicted DNA sequence differences in the 5' and upstream *GLI* sequences.

This problem of potentially parallel expression and function of the human GLI2 and GLI3 proteins was considered critical enough to include a coarse analysis of the *GLI2* 5'-region in this thesis in parallel with the more detailed study of *GLI3* regulation.

6.1 GLI3 morphopathies without mutations in transcribed sequences are candidates for defects in regulatory sequences

To search for patients with GLI3 morphopathies not associated with mutations in coding sequences, 24 patients were screened for changes affecting sites of the GLI3 protein that are essential for its function. Such samples are very useful candidates for mutation screening in elements residing upstream, downstream or in the intronic regions of *GLI3* that may play an important transcriptional regulatory role. Previously, in one Greig syndrome patient with a translocation dissecting chromosome 7 about 10 kb downstream of *GLI3*, a positional effect involving regulatory elements was predicted (Vortkamp et al. 1991). In the mouse, the anterior digit deformity phenotype (add) is caused by a transgene integration ~40 kb upstream of *Gli3* which causes the expression of this gene to decrease (Schimmang et al. 1993; van der Hoeven et al. 1993). Similar observations were reported for other developmental genes. For instance, the *Ssq* mouse, that shows pre-axial polydactyly (PPD), is caused by a mutation in a conserved region inside the *Lmbr1* gene which affects the correct spatial expression of the distantly located gene sonic hedgehog (*Shh*), coding for the signaling factor upstream of GLI3 (Lettice et al. 2002).

The patients studied in this project for point mutations in their coding exons, previously, had been analyzed for cytogenetic alterations. No abnormalities had been detected. Among the 24 cases four showed point mutations. Two of those had been detected previously in unrelated individuals: A missense mutation in the 13th exon (c.1873 C>T; R625W) was diagnosed in a GCPS patient with syndactyly in all the limbs, polysyndactyly in the left and right foot, high forehead and bregma, frontal bossing, hypertelorism and broad nasal root. A C is exchanged by a T in the 1873rd nt of the ORF, which results in the substitution of an arginine with a tryptophan. The mutation altered the 625th amino acid of GLI3 within the zinc finger domain. This mutation was transmitted to the daughter of the patient, which shows the same clinical features. In both cases, the mutation is heterozygous. The second previously reported missense mutation in the 15th exon (c.3664 C>T; P1222S) was identified in another GCPS patient with hexadactyly. A C is exchanged by a T in the 3664th nt of the ORF, which results in a substitution of a proline with a serine. The mutation altered the 1222nd amino acid within the transactivation domain 1, TA1. The mother of this patient is heterozygous for the mutation but does not show any phenotype whereas the father's coding sequence is unaffected.

In the first case, the amino acid exchange from R to W occurred in a region highly conserved between human, mouse, dog and zebrafish protein sequences, with the R present at this position in all the species (CAB59315 for human, NP 032156 for mouse, XP 540363 for dog (predicted) and NP 991291 for zebrafish GLI3).

The high conservation of this site suggests that it is important for the protein function. Since this mutation is located in the zinc finger domain it is possible that it alters the DNA binding ability of GLI3. In the second case, the exchange of the P to an S occurred at a site that is conserved between human and mouse, however, a corresponding sequence cannot be found in the dog or the zebrafish. Possibly, this mutation affects a site that may be less essential for the function of the GLI3 protein as evolution did not conserve this part of the protein between more distantly related species. The region where this mutation occurred is in the transactivation domain 1, and it is possible that a minor variation in protein sequence does not affect the function of this domain of GLI3. It must also be noted that the mutation in the zinc finger region is associated with the phenotype in the family, whereas the second mutation is not strictly correlated with the phenotype. The patient's mother is reported as unaffected although she carries this mutation.

In a previous, unrelated case, a missense mutation of the 808th amino acid exchanged an I with an M, residing in a region for which no functional domain is predicted, although the I is conserved between the human, mouse and dog. This mutation was also found in the phenotypically unaffected mother of this patient. Therefore, this amino acid exchange seemed not to affect the function of GLI3.

Both this variant and the one detected in this thesis do not occur as a polymorphism in the general population. It is possible that we are detecting in these cases very rare mutations, which are not functionally associated with the limb phenotype. Alternatively, they may represent mutations predisposing to GLI3 morphopathies if combined with a second, unidentified GLI3 mutation or a change in a different, functionally interacting factor.

The first novel mutation that has been identified in this work affects two base pairs altering two neighbouring codons. This generates a double mutation, one silent and one nonsense in the 13th exon (c.2058_2059 GG > AT; R686R, E687X). The truncating mutation at the 2059th nucleotide of the GLI3 open reading frame disrupts the protein in the 687th amino acid, immediately after the zinc finger domain. Mutations in this region have been found to cause mostly Pallister-Hall syndrome (PHS) (Johnston et al. 2005). Indeed, this patient features mild syndactyly, postaxial polydactyly and broad terminal phalanges in both hands, hypertelorism, broad nasal root, short structure, growth hormone deficiency and hypothalamic hamartoblastoma, characteristic for PHS. This mutation has occurred de novo, as it could not be identified in the parents. The observation of this novel mutation supports the notion of a genotype-phenotype correlation of PHS with truncations of GLI3 leaving zinc finger region and the putative N-terminal repressive region intact while eliminating the rest of the protein (Johnston et al. 2005). The silent mutation, which does not change the amino-acid sequence of the protein, most probably does not play any role in the creation of the GCPS phenotype.

The second novel mutation (c.4675 A>G; I1559V) is a missense one in the 15th exon, and the patient shows an acrocallosal syndrome (ACS). An A is exchanged with a G in the 4675th nt of the ORF, which results in the substitution of an isoleucine with a glycine in the 1559th amino acid in the transactivation domain 2, TA2. This amino acid and the protein sequence of this region are highly conserved between human, mouse, dog, and zebrafish. Thus, this part of the protein can be predicted to have an essential function. The sequence of TA2 is more conserved between distant species

than the sequence of TA1, which was mutated in a case discussed above. Perhaps, TA2 is functional in GLI3 proteins of distant species whereas the role of TA1 is restricted to a closer range of related species like human and mouse.

The parents of the patient are phenotypically normal and appear not to be carriers for this mutation, i.e. this mutation most likely has occurred de novo.

The detection of this mutation associated with an acrocallosal syndrome is noteworthy because this phenotype, previously, has been reported not to be linked to GLI3 (Brueton et al. 1992). Only recently, a case of a child with agenesis of the corpus callosum and severe retardation, both cardinal features of ACS and rare in GCPS, has been reported, who similarly has a mutation in GLI3 (Elson et al. 2002). To clarify this issue, functional studies will have to prove in each case of ACS that the mutation changes the function of GLI3 and could be causal for the phenotype. The repeated, although rare, observation of GLI3 mutations in ACS could indicate genetic heterogeneity of this phenotype.

The majority of potential GLI3 morphopathies studied for this work did not reveal mutation in coding sequences that could explain the phenotypes. They are valuable candidates for mutation searches in regulatory elements.

6.2 Human GLI2 is lacking DNA sequence homology with GLI3 in the 5' region and the promoter

Both GLI2 and GLI3 from human and mouse code for proteins with an extended N-terminal domain, followed by the zinc finger domain, and ending with the transactivation domain at the C-terminus (<http://genome.ucsc.edu/>), (Roessler et al. 2005), in contrast to the human and murine GLI1 that does not possess most of this part. However, no functional promoter element or an initiator site has been yet identified for human *GLI2*. In order to understand the transcriptional regulation of this gene, and in view of the fact that *GLI2* and *GLI3* show overlapping and distinct expression patterns, a search for an unreported transcribed sequence upstream of the previously published 1st exon of *GLI2* was initiated at the RNA level.

For RT-PCR, placenta cDNA was chosen as a starting template to perform 5' RACE PCR starting from the second published *GLI2* exon. The PCR product contained, in addition to the known exonic sequence, a 1.2 kb sequence element that is identical to the published genomic sequence upstream of the first *GLI2* exon. This sequence,

which is annotated as non-RNA, could be amplified in one fragment with the 1st and 2nd exon of *GLI2*. That proves that this 1.2 kb region is part of the cDNA together with the already annotated exons (figure 6-1). It remains to be elucidated whether the RT-PCR identified the definite start of transcription or if more upstream sequence exists in the *GLI2* cDNA in other tissues.

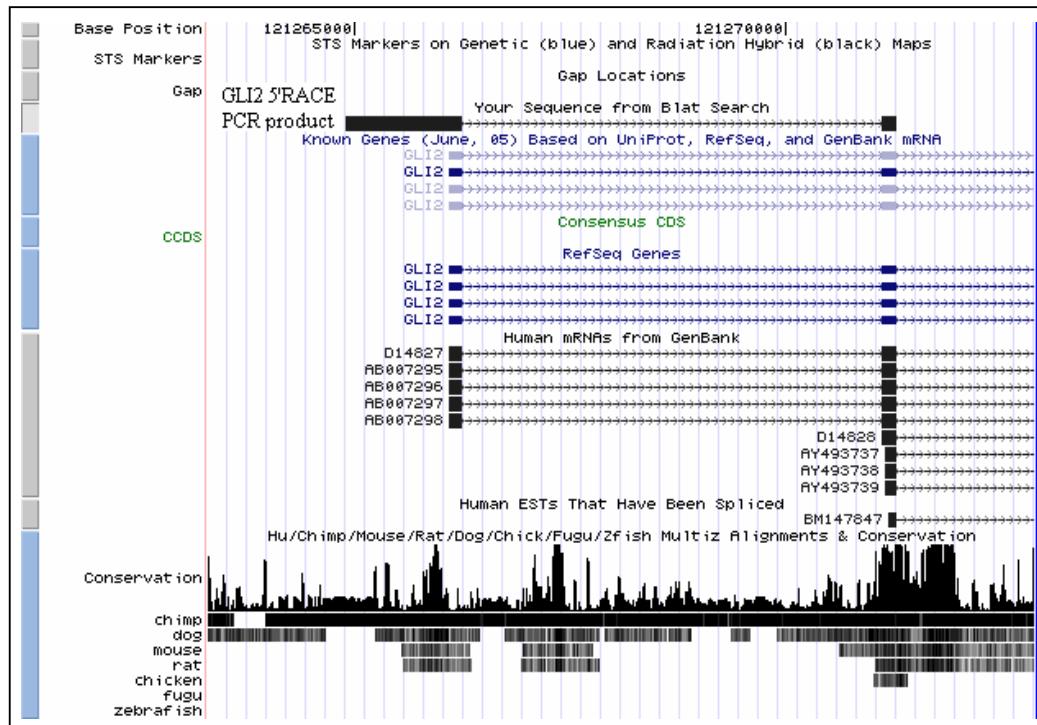


Figure 6-1: Location of the human *GLI2* 5'RACE product

The position of the *GLI2* 5' RACE PCR product is depicted in comparison to the annotated *GLI2* sequence of the UCSC Genome Bioinformatics site. The similarity starts with the 2nd *GLI2* exon, where the *GLI2* specific reverse primers annealed, continues upstream into the 1st exon and extends approximately 1.2 kb further upstream. Human mRNAs from GenBank as well as an intraspecies comparison are also shown.

In addition, conserved regions are highlighted in the exonic as well as in the intronic regions. Conserved intronic sequences cannot be found in the fish but only in more related species such as chimp, mouse, rat and chicken. Further experiments can show whether these homologies contain regulatory elements or are in any other way functional.

The newly identified *GLI2* cDNA sequence as well as a 1 kb upstream genomic fragment was searched *in silico* for potent promoter elements. The predicted promoter elements upstream of the human and mouse *GLI2* genes differ in their sequence, their position and their characteristics (figure 6-2).

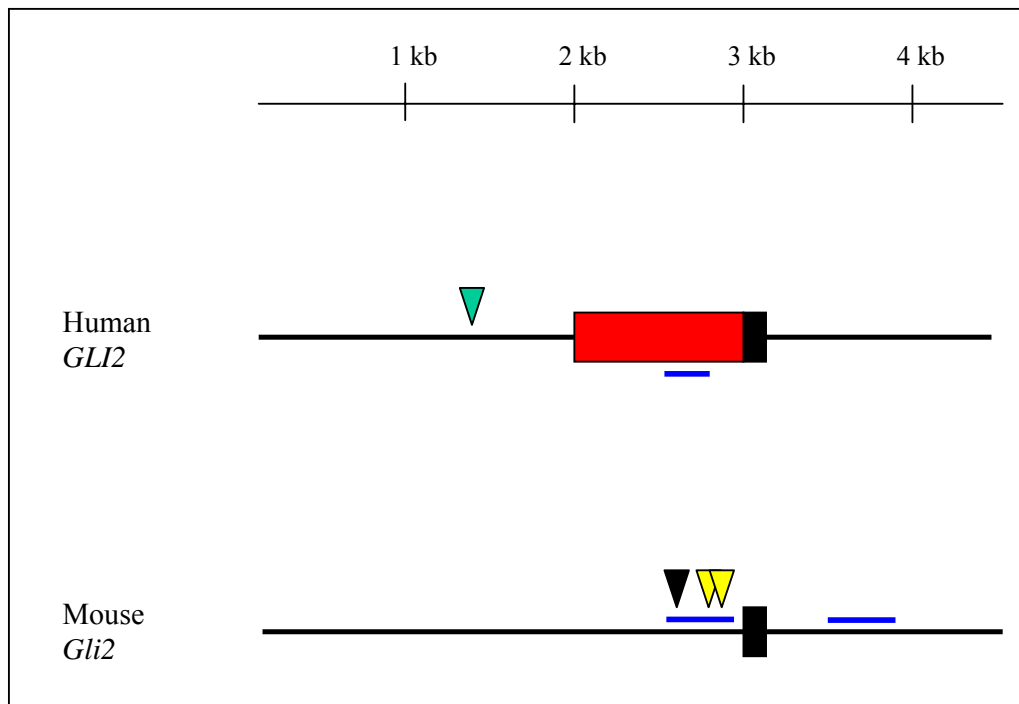


Figure 6-2: Location of predicted promoter elements in human and mouse *Gli2* upstream sequences

Genomic DNA 2 kb upstream and 1 kb downstream of the published 1st exon of human and mouse *GLI2* were scanned by making use of the NIX analysis resource of the HGMP Resource Centre, Hingston, UK, (<http://www.hgmp.mrc.ac.uk/>) with the “GRAIL-CpG” program and with “TSSW” to predict promoters and intron sites. The bold black lines represent genomic DNA, the boxes depict the 1st exons of both genes and the red box is the expanded part of the *GLI2* 1st exon using the 5’ RACE PCR. TATA-less and TATA-containing promoters are shown with black and green triangles, respectively, intron sites with yellow triangles and CpG islands with blue lines.

The human *GLI2* has one 208 bp long predicted CpG island that resides inside the newly identified sequence and a predicted promoter element containing a TATA box that resides 682 bp upstream of the site where the 5’ RACE PCR stopped. CpG islands are regions of a high CpG dinucleotide content and are regarded as useful landmarks in the genome for the presence of genes (Larsen et al. 1992). CpG islands can start upstream of the 1st exon of a gene and stretch into the 1st intron, as is the case for the *GLI3* gene (Vortkamp et al. 1995). The functionality of the predicted TATA box containing promoter element has to be experimentally confirmed. Both *GLI1* and *GLI3* genes are regulated by TATA-less promoters, therefore it would be an exception

if *GLI2* transcription would be driven by a TATA-containing promoter. The murine *Gli2* contains two putative CpG islands. The first is 331 bp long ending 93 bp upstream of the 1st exon and contains two predicted inr sites and a predicted TATA-less promoter element. These elements still have to be experimentally verified for their function. The second CpG island is 452 bp long and resides 382 bp downstream of the 1st exon. Comparing the genomic DNA sequence of these 3 kb regions (2 kb upstream and 1 kb downstream of the 1st exon of each gene) between human and mouse *Gli2*, no significant similarity can be found.

The upstream DNA sequences of the *Gli3* genes share to some extent homologies between human and mouse: Two homologous regions can be found, one 40 bp long and of 70 % identity located just upstream of the potential initiator site determined for the human gene with the primer extension method, and another one 52 bp long and of 77 % identity further upstream that resides inside a trinucleotide repeat ([GGA]₇ GAA [GGA]₆ [GGC]₉). The cDNA start of the murine *Gli3* is located 97 bp further upstream of this second homology (Jacobsen 1996), and its sequence does not resemble the cDNA starts of human *GLI3* identified with either method. In addition, a 1.5 kb CpG island is found in the upstream region of the murine *Gli3*, including the promoter region and reaching until inside the 1st intron (figure 6-3).

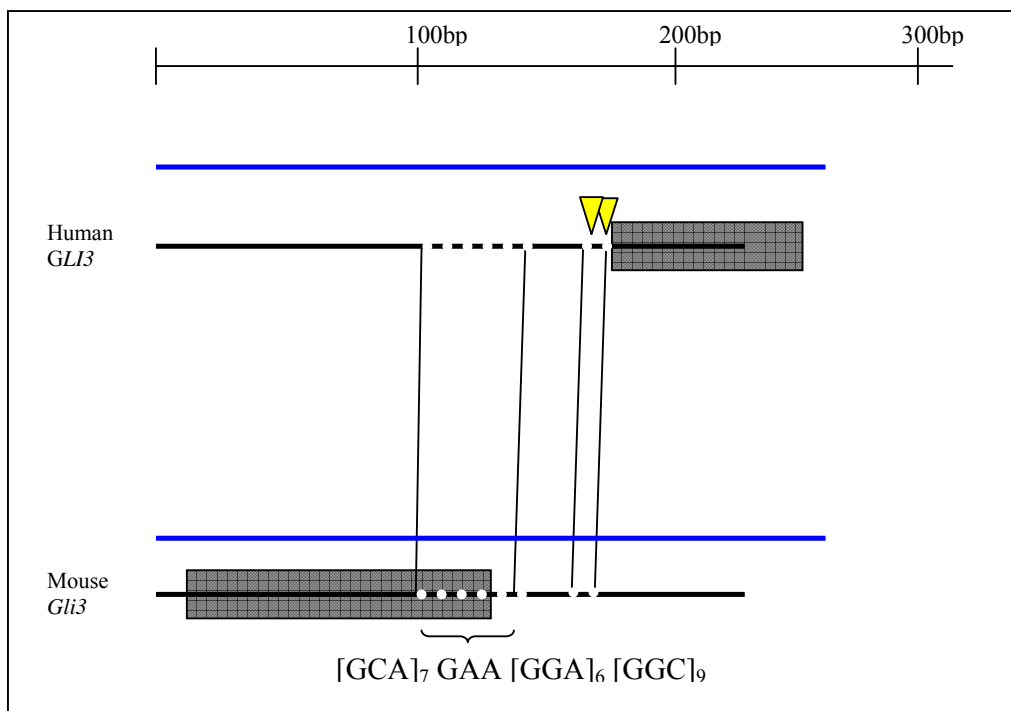


Figure 6-3: Comparison of the promoter regions between human and murine *GLI3*.

The bold black lines represent genomic DNA, the black dots are the conserved regions between the two promoters. The black semitransparent boxes are the 1st exons of both genes and the yellow triangles depict the experimentally identified *inr* sites. The CpG islands are depicted with blue lines; their complete length is not shown in this figure.

The differences between the upstream regions of human and mouse *Gli3* as well as between the *Gli2* genes of these species suggest that during evolution non-coding but regulatory important sequence elements of the GLI genes were not conserved in spite of their similarity in expression patterns.

The additional cDNA sequence amplified from placenta cDNA was checked also for the presence of open reading frames. However, in none of the three frames an ORF could be found that is compatible with the annotated GLI2 sequence.

Such a large, untranslated cDNA fragment upstream of the 1st published *GLI2* exon can be explained by a chromosomal rearrangement such as a translocation or a deletion of a DNA fragment in this region specific for the evolution of primates. This event might have attributed to GLI2 novel functions or might have restricted its potential to regulate target genes, thus playing a role in the specificities of the human phenotype.

6.3 Analysis of the *GLI3* minimal promoter

A 3.3 kb region upstream of the predicted start of the *GLI3* 1st exon had previously been deleted successively. The fragments were tested for their potential to drive expression of the luciferase reporter gene in transient transfection cell assays. By this way, a fragment containing 300 bp upstream of the 1st exon was identified as the minimal promoter of *GLI3* (Jacobsen 1996). Unfortunately, the previously generated constructs were no longer available for this study. To avoid ambiguities originating from different transcription start sites, two further promoter fragments corresponding to the previously defined “-600 bp *SacI* 5’ deletion” and to the “-300 bp *ApaI* 5’ deletion” constructs but reaching further into exon 1 were now generated and used throughout in this project (figure 6-4). The construct fragment “promGLI3-300” contained the previously defined minimal promoter.

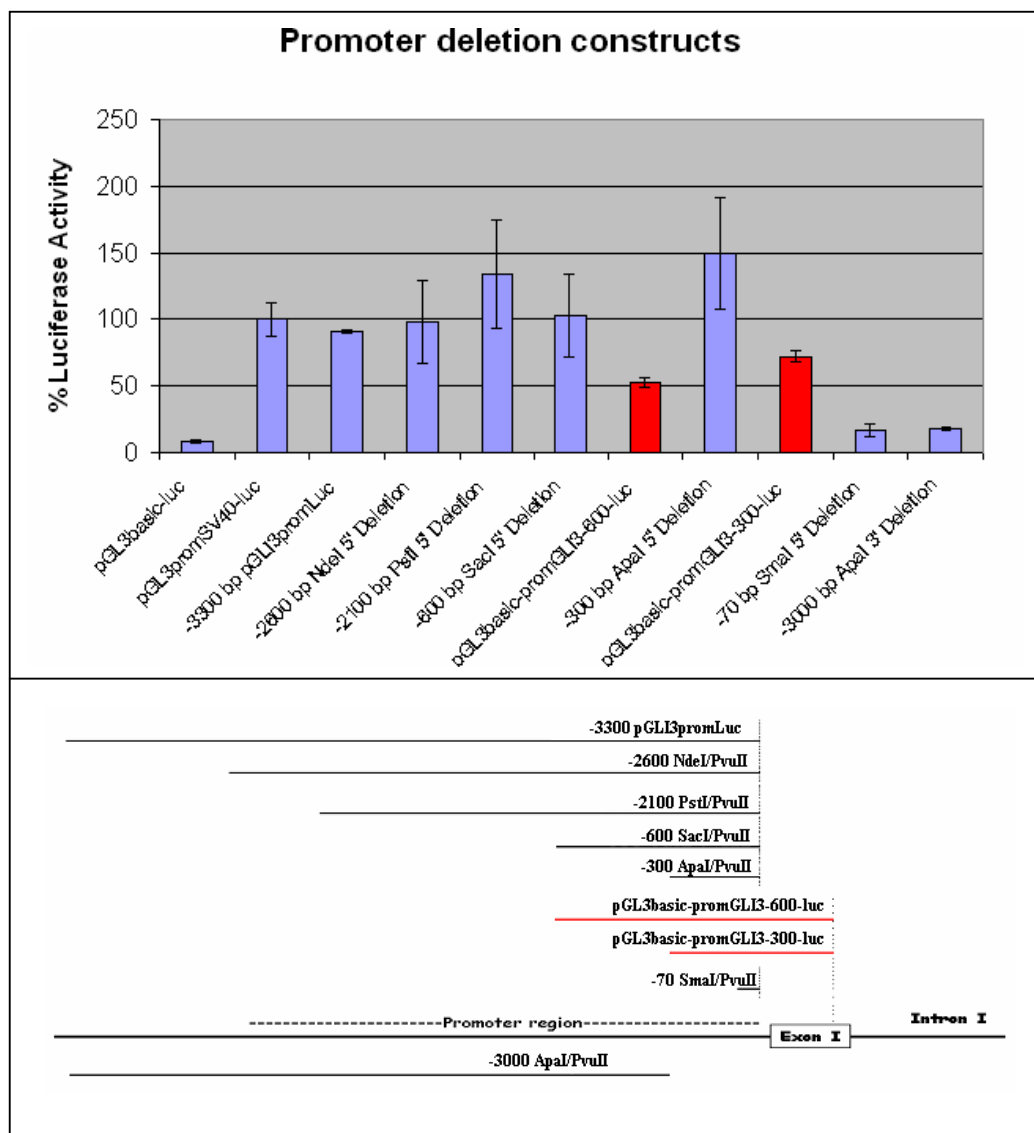


Figure 6-4: Luciferase expression achieved using different promoter deletion constructs

Luciferase activity obtained in H661 cells transiently transfected with constructs expressing the gene under the control of human genomic fragments extending from *GLI3* exon I upstream (upper part). The length of the promoter fragments and their name is indicated in the bottom part. Results from Jacobsen (1996; bottom part black lines, upper part blue bars) are integrated with recent results (red lines and bars).

When compared to the activity of an SV40 promoter construct, the activity of the *GLI3* promoter including part of the 1st exon of *GLI3* appears to be decreased relative to the -300 bp Apal/PvuII construct from (Jacobsen 1996), which ended 61 bp

upstream of the old cDNA sequence. This might be attributed to the additional sequence cloned in the pGL3basic-promGLI3-300-luc construct.

However, since the original constructs were unavailable for comparison the difference might likewise be attributed to different experimental conditions. This notion is corroborated by a similarly decreased activity of the pGL3basic-promGLI3-600-luc construct compared to the previous measurements (figure 6-4).

To eliminate the ambiguities originating from the lack of an experimental definition of the transcription start site it was decided to determine the initiator (inr) site. Interestingly, the primer extension method and the 5' RACE PCR identified two different sites where mRNA synthesis of *GLI3* starts. According to the primer extension cDNA starts 123 nt upstream of the ATG within exon 2 with which the ORF of *GLI3* starts, whereas after the 5' RACE PCR the longest transcript was found to start 133 nt upstream of this site, and exactly at the start site position predicted by the "TSSW" program. The site that was used to clone the promoter fragments by Jacobsen 1996 resides 2 nt downstream of the site where the 5' RACE PCR terminates. Therefore, these previously used fragments can be considered to contain the complete *GLI3* promoter.

Human skeletal muscle cDNA was used for the primer extension whereas placenta cDNA was analysed by the 5' RACE PCR. In these tissues, the *GLI3* gene may use different inr sites to start its transcription. Genes using two or more inr sites are not unusual such as in the promoter of Emilin1 (Fabbro et al. 2005). The transcription start site according to the primer extension overlaps with another predicted inr site that follows the consensus py-c-a-n-t-py-py (py stands for pyrimidine) (Javahery et al. 1994). The C and the A of this consensus sequence were each exchanged with a T. The "pGL3basic-promGLI3-300-luc" construct carrying the *GLI3* minimal promoter with the initiator site containing the two mutated nucleotides was transfected into the H661 cell line but the mutation did not show any effect on the promoter activity. Possibly, the second initiator site is used in this cell line. Therefore the mutation does not influence the activity of the promoter. The extent to which each inr site is used for the transcription of *GLI3* still has to be determined also in other cell lines. No downstream promoter element (DPE) can be identified downstream of either inr site. No TATA-box or a CAAT-motif could be identified in the *GLI3* promoter. TATA-less promoters usually belong to genes that encode "housekeeping" enzymes, oncogenes, growth factors and their receptors, and transcription factors. For the

transcription of such genes, the E2F and the YY1 transcription factors have been reported to play a major role (Azizkhan et al. 1993). However, binding sites for either element could not be identified in the *GLI3* minimal promoter.

Individual mutation of five predicted transcription factor binding sites within the pGL3basic-prom*GLI3*-300-luc construct reduced the activity of the promoter only slightly, if at all (figure 5-13). These sequences contain sites for TFs that are usually involved in promoter regulation, such as Sp1, HiNF-C, c-Ets2, CREB, AP-1, MAZ and CACCC-binding factors (Xie et al. 2005). In contrast, the further truncation of the promoter by 126bp caused the decrease of its activity by half (figure 5-13). Obviously, sites contained in the deleted sequence are important for the activity of the *GLI3* promoter, but due to the fact that most of these TFs bind in many different sites on the promoter, a single mutation in one site cannot infer a sizeable decrease in the activity of the construct. The residual activity of the truncated promoter can be attributed to the existence of further TF binding sites in the remaining sequence. A 70 bp long fragment residing just upstream to the newly identified start of the *GLI3* cDNA had shown very low activity in transient transfection assays (Jacobsen 1996). Therefore, the remaining promoter activity after the truncation of the minimal promoter resides in a 142 bp fragment situated between the truncation site and the start of the 70bp fragment that showed very low activity. Binding sites for SP1, PPUR, CUP, EGR2 and WT1-KTS factors dominate in this 142 bp promoter fragment. Since its sequence is abundant in G/C, it was impracticable to design primers for site directed mutagenesis of these binding sites in order to test their involvement in the regulation of *GLI3* expression. In addition, these factors could bind repeatedly along this sequence. Hence further deletion of small fragments might be the best strategy to elucidate the regulation of the *GLI3* minimal promoter.

The activity of the *GLI3* minimal promoter that is contained in the pGL3basic-prom*GLI3*-300-luc construct in the different cell lines is correlated to their endogenous *GLI3* expression. When transfected into the *GLI3* expressing H661 and U2OS cell lines, the minimal promoter shows a high activity. These cell lines express endogenously necessary factors for the *GLI3* expression. The measurements did not reveal a significant difference in the potential to promote reporter gene expression in these two cell lines (figure 5-11). The H441 line, which does not express *GLI3*, showed a very low activity of the pGL3basic-prom*GLI3*-300-luc construct. The lack

of elements in these cells, which are necessary for the endogenous *GLI3* expression, appears to affect the ability of the transfected construct to express the luciferase reporter gene.

6.4 A region in the 5' part of intron 1 enhances the action of the *GLI3* minimal promoter

Two constructs were studied which contain the same upstream region of the *GLI3* promoter but extend to different lengths into intron 1 (figure 5-16). The larger one (pGL3basic-3.7kbStuI-StuI-luc) includes a region which is highly conserved between human and mouse DNA and is situated 2-2.2 kb downstream of exon 1 whereas the shorter one ends 500bp downstream of the 3' end of the 1st exon.

The shorter one (pGL3basic-1.9kbStuI-DpnI-luc) shows a very high activity up to 400% of the minimal *GLI3* promoter activity (320% of the SV40 promoter activity), whereas the larger construct shows just 63% of the minimal *GLI3* promoter activity (50% of the SV40 promoter activity) (figure 5-17). The very high potential of the short construct to drive expression of the luciferase gene, in comparison to the activity of the minimal promoter, may be attributed to the existence of TF binding sites downstream of the 1st exon, which are necessary for the up-regulation of the *GLI3* minimal promoter. These factors can be contained in short sequences that may be conserved between human and mouse but they are probably too short to be identified with the VISTA program.

The “TSSW” program predicts promoters and transcription start sites in this intronic region downstream of the 1st exon. This could suggest that the expression of *GLI3* can be regulated by an alternative promoter, which resides or expands until inside the 1st intron. An alternatively spliced 90 bp-long exon of *GLI3* is predicted to start 724 bp downstream of the 3' end of the 1st exon, according to the UCSC Genome Browser database. The expansion of the CpG island up to 722 bp downstream of the 1st exon as well as the prediction of a good promoter and initiator site 80 bp upstream of this predicted exon offers the possibility that an alternative promoter would be able to drive expression of a transcript that would give full length, functional *GLI3* protein as the ORF starts from the 2nd exon (figure 6-5). This predicted alternative promoter was not tested for its activity in this work.

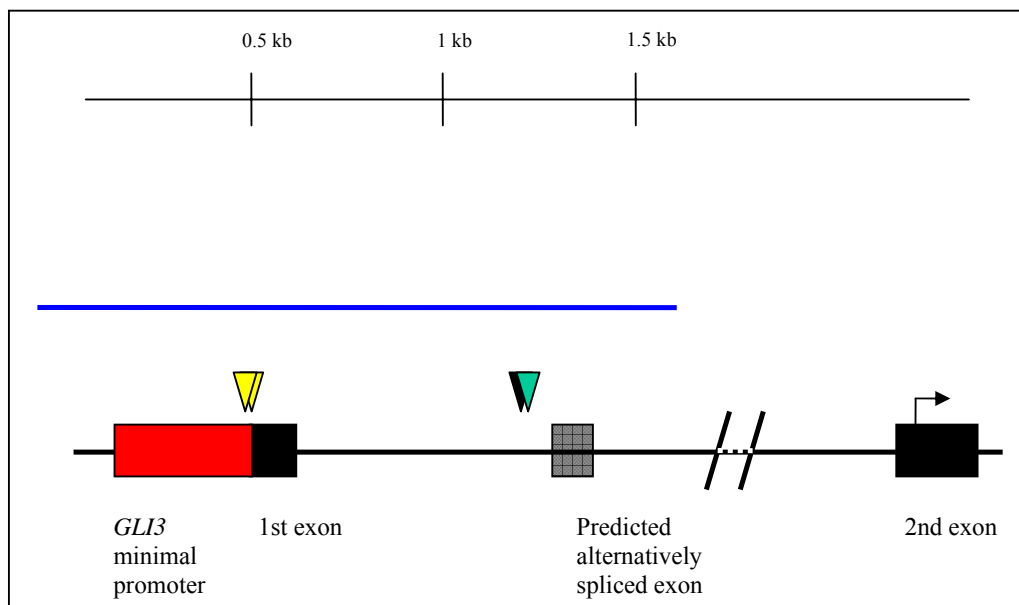


Figure 6-5: Position of predicted alternatively spliced promoter.

The bold black line represents genomic DNA. The black boxes depict exons whereas the semitransparent one is the predicted alternatively spliced exon. The yellow triangles are the experimentally identified intron sites, the green and black triangle the predicted intron site and promoter respectively. The arrow depicts the start of the *GLI3* ORF. The CpG island is drawn with blue lines, and its complete length is not shown in this figure.

The low activity of the larger construct may be explained by the existence of down-regulators in the extra sequence. An interesting candidate for this function might be the highly conserved sequence between human and mouse DNA that resides 2 – 2.2 kb downstream of the 1st *GLI3* exon (figure 6-6).

experiments this fragment showed a very high potential of driving reporter gene expression (~140% compared to the SV40 promoter) without the addition of a promoter when cloned in the positive orientation, but almost no activity when negatively oriented. In further series of experiments the activity of this construct reached only 40% of the activity of the SV40 promoter although seemingly no experimental variable had been altered. Even with the decreased activity, Hom2 shows a 4 times higher potential to drive expression of the luciferase gene than the promoterless pGL3basic-luc construct.

To test, whether Hom2 contains promoter elements and the transcription of the luciferase gene starts inside the Hom2 sequence when transfected in the H661 cell line, primer extension was used to identify the transcription initiation site of this construct. The primer extension analysis on total RNA extracted from H661 cells transfected with the pGL3basic-Hom2(+)-luc construct resulted in three bands. (figure 5-27). Unfortunately, the sequencing results are insufficient in the critical interval, however, it is evident that the transcription from this construct starts from a region ~40 bp upstream of the 3' end of the Hom2 sequence. In the 13 nt of the Hom2 to which the primer extension bands correspond, only one weak inr site can be predicted that does not match absolutely the inr site consensus sequence.

Inr consensus sequence : py-c-a-n-t-py-py

Predicted in Hom2: t-**g**-a-a-t-t-**g**

The red, bold letters are the nucleotides from the Hom2 sequence that do not match the consensus of an inr site. It is difficult to characterize this sequence as a promoter as there are three different, closely situated, possible transcription initiator sites and no other promoter elements predicted in it. In addition, this homology is identified in the 2nd intron, ~4.5kb upstream of the 3rd exon. Another alternatively spliced *GLI3* can be identified to start 330 bp downstream of the 3' end of Hom2. This exon is predicted in the AceView database, which is developed by NCBI. Such alternative exons originate from cDNAs that have been submitted to the GenBank and have been aligned with the genomic DNA. Interestingly, the conservancy of this region between human and mouse expands up to the beginning of the 1st exon of this alternative transcript. This alternatively spliced gene would start from the 3rd exon of the original *GLI3*, thus missing part of its upstream repressor domain of the protein (figure 6-7).

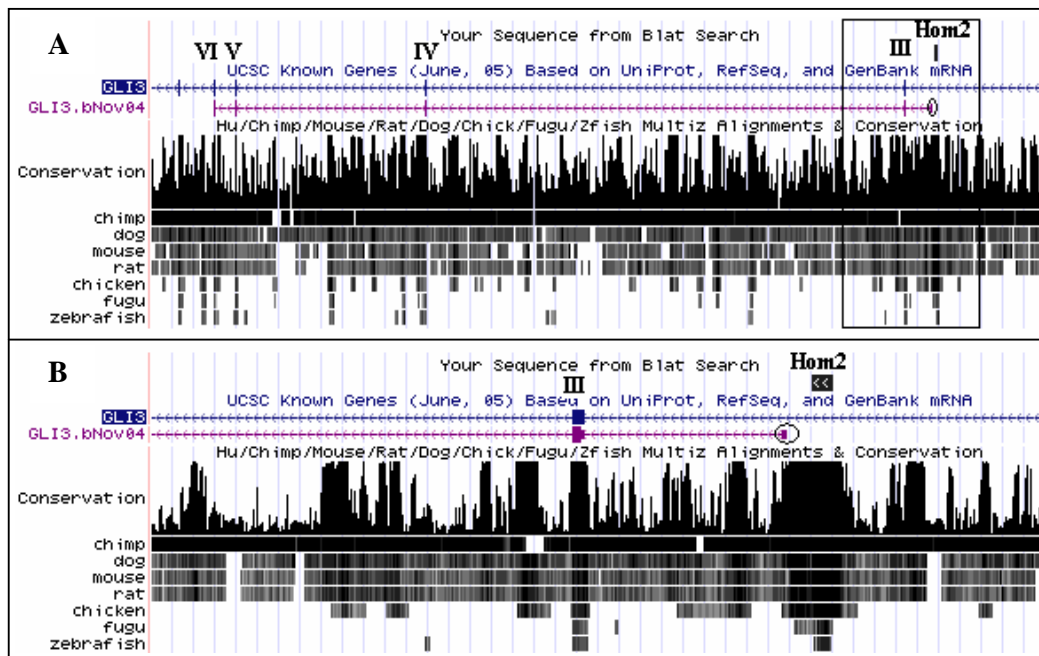


Figure 6-7: Location of Hom2 within the *GLI3* gene

A: The position of the Hom2 is depicted in comparison to the annotated *GLI3* sequence (blue line) of the UCSC Genome Bioinformatics site and to a predicted alternatively spliced *GLI3* model (mauve line). The sequence conservation between different species is also shown. **B:** The region in the rectangle in **A** is shown with more detail. The difference in the length of the conservation (thick black line) of this region between human and mouse and between human and fugu can be seen clearly. The conservancy between human and mouse stops just upstream of the alternatively spliced exon. The direction of transcription is from right to left, and the exons are shown with Latin numerals. The predicted alternatively spliced exon is circled.

However, no other promoter elements could be predicted in Hom2 or in the intronic sequence between this region and this alternative exon.

To identify a transcription factor that might interact with Hom2 driving reporter gene expression, binding sites predicted by the TESS program were inactivated by mutagenesis (figure 5-23). The three mutations of the Hom2 sequence in different sites, creating each time a new BamHI restriction site, resulted in the same decrease of the pGL3basic-Hom2(+)-luc activity (figure 5-24). A TF binding site that is shared by all the mutated sequences is NFATp. The nuclear factors of activated T cells (NFAT) proteins are a family of TFs, which are activated by calcineurin. They have been first

identified as inducers of cytokine gene expression in T cells but they play also various roles in other tissues. NFATp participates in the prostaglandin synthesis during angiogenesis, in skeletal muscle growth, in inhibition of chondrogenesis, in expression of adipocyte specific genes during differentiation, in embryonic axonal outgrowth, in the heart valve development and in the myf-5 expression (Horsley and Pavlath 2002; Graef et al. 2003). Myf-5 is also a downstream target of GLI3 (McDermott et al. 2005). Additionally, in order for NFAT proteins to function, they must form homo- or heterodimers with FOS and JUN (McCaffrey et al. 1993) and binding sites for these TFs exist in the sequence of Hom2.

The identification of binding transcription factors to a given site depends on their presence in the cells or tissues to be tested. The potential of Hom2 to promote reporter gene expression, therefore, has also been tested in other cell lines. In the U2OS cell line, no NFATp expression could be identified with RT PCR analysis in contrast to the other two cell lines (figure 5-26). In the GLI3 expressing U2OS cells the activity reached 40% of the SV40 promoter activity whereas in the GLI3 negative H441 cells its activity was completely diminished (figure 5-25) despite the expression of NFATp. Interestingly, the activity of wild type Hom2 in USOS was almost the same as the activity of mutated Hom2 constructs in the NFATp expressing H661 cell line (figure 5-24). Perhaps NFATp needs all of the three binding sites on this region in order to regulate the activity of Hom2. This observation is compatible with an involvement of NFATp in Hom2 regulation.

There are other candidates for transcription factors that might bind on Hom2. For instance, AML1 (RUNX1) and its family members (RUNX) are factors that regulate the expression of genes involved in cellular differentiation and cell cycle progression. RUNX1 has repression and gene silencing activities (Durst and Hiebert 2004), thus such a protein may strongly repress the activity of Hom2 in the H441 cell line. It remains to be determined whether this protein is present in the H441 cell line and is absent from the other two.

As will be discussed below, Hom2 acts as a tissue specific enhancer of a reporter gene in transgenic mice. The observation that Hom2 cannot up-regulate the *GLI3* minimal promoter in the cell assays might be explained by assuming that this promoter already has reached its maximum activity in this cell line and cannot be further enhanced. An

enhancing function of Hom2 may be required in tissues where *GLI3* promoter is less active and needs to be up regulated in order to express this gene. It is also possible that Hom2 interacts with a site of the *GLI3* promoter, which is not present in the construct used for the cell assays.

The 2nd analysed highly conserved non-coding region (Hom3) resides 23 kb downstream of the 3rd exon of human *GLI3*, inside the 3rd intron. It shows approximately 98% homology between human and mouse and around 75% homology between human and fugu DNA. Transient transfection assays showed a negative regulatory activity of this fragment. When this fragment is cloned upstream of the human *GLI3* minimal promoter driving the luciferase gene, it decreases the activity of the promoter from 80% to 40%, relative to the activity of the SV40 promoter, regardless of its orientation (figure 5-22). This region was truncated in two overlapping fragments. One contains the 5' sequence and the other the 3' sequence of Hom3, both share the central part. The activity of these constructs reached the same level as the whole Hom3 (around 40% of the activity of the SV40 promoter, figure 5-30), thus the elements that are responsible for the down-regulation of the *GLI3* promoter are expected in the central part of the Hom3 sequence which is shared by the two truncated fragments. These observations were obtained with constructs in either orientation. The only exception is the negatively orientated fragment that includes the 5' region of the Hom3. The observed repression of the promoter in this case is stronger and the activity of this construct decreases to 20% of the activity of the SV40 promoter (figure 5-30). There are reports of cis-regulatory elements that have different activities depending on their orientation or which act only in a specific direction (Lin et al. 2004; Emoto et al. 2005). Thus, the fragment of the Hom3 missing its 3' part may act also differently according to its orientation.

Within the sequence of the Hom3 binding sites for factors can be identified that might have a negative regulatory role. One of them is E4BP4, which contains a C' terminal repressory domain and is expressed in the lung, the tissue H661 is derived from (Cowell et al. 1992; Hulme et al. 2000). Another one is GATA1 which is a negative regulator of gamma-globin in adults and interacts with its promoter in association with Sp1 or CCACC binding proteins (Walters and Martin 1992; Eleouet and Romeo 1993; Fischer et al. 1993). However, GATA1 is an erythroid-specific transcription factor that is expressed in the haematopoietic and the Sertoli cells in mice (Whitelaw

et al. 1990; Walther et al. 1996). Therefore, it is difficult to assume that it can play a role in *GLI3* regulation. Hom3 by itself could not drive expression of the luciferase reporter by itself gene when cloned in either orientation upstream of this gene in a construct without promoter.

Like Hom3, also the third analysed highly conserved non-coding sequence element, Hom4, showed the ability to reduce the *GLI3* minimal promoter activity in both orientations (figure 5-22). Hom4 resides 3 kb downstream of the 4th exon, inside the 4th intron of human *GLI3*. The negative effect is not as strong as the effect of Hom3, but still, when Hom4 is cloned upstream of the minimal *GLI3* promoter in both orientations, reporter activity falls down to 60% of the SV40 promoter. Negative regulatory TFs can be predicted to bind in Hom4 (figure 6-8).

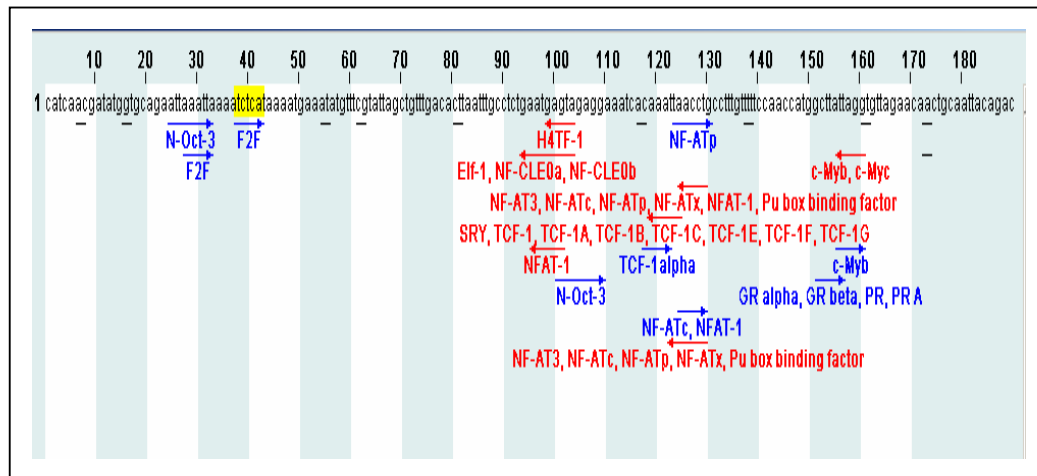


Figure 6-8: TESS plot for the Hom4 sequence

The arrows indicate predicted binding sites for various TFs in the Hom4 sequence.

One of them is F2F which is expressed ubiquitously and is most likely a transcriptional repressor (Jackson et al. 1992). Another one, the NF-CLE0 contains sequence with inhibitory activity (Miyatake et al. 1991). Interestingly, there are also NFATp binding sites predicted, for which a positive regulatory role was postulated upon Hom2. However, in order for NFAT proteins to function, they must form homo- or heterodimers with FOS and JUN, which are not predicted to bind in the Hom4 sequence, in contrast to Hom2.

Homology searches did not identify elements corresponding to Hom2, Hom3, or Hom4 in the sequences of *GLI2* or *GLI1* from either human or mouse (data not

shown). Thus, apparently no obvious paralogs of these highly conserved sequence elements exist within the non-coding parts of the paralog *GLI* genes. If the analysed elements are involved in directing tissue specific gene expression of *GLI3*, it remains to be demonstrated, how the coordinated expression of the paralog *GLI* genes is achieved.

6.6 Hom2 but not Hom3 and Hom4 direct tissue specific expression of a reporter gene in transgenic mice

The three highly conserved non-coding DNA regions from *GLI3* introns under study were cloned in front of the human beta-globin promoter adjacent to the lacZ reporter gene. In transgenic mice, the beta-globin promoter apparently does not drive expression of the reporter gene by itself (Simmons et al. 2001). By cloning an enhancer element in front of this promoter, the expression is activated, potentially exactly in the tissues in which the cis-regulatory element is responsible for switching on its target genes.

Hom3 and Hom4 were inserted in both orientations in appropriate reporter constructs and used to generate transgenic mice. None of them induced any lacZ staining in 9 or 11 days old mouse embryos, although several embryos were identified by PCR analysis to contain the injected lacZ gene in their DNA. This result corroborates the observations seen in transiently transfected cell cultures in which these fragments showed negative regulation of the minimal *GLI3* promoter. As the beta-globin promoter by itself does not activate transcription, adding an element with negative regulatory properties should not activate expression of lacZ in any tissue.

Hom2 was injected only in the positive orientation. One out of twelve 9-day embryos exhibited whole mount beta-galactosidase staining restricted to specific tissues, in particular in the developing forebrain (figure 5-32). Three out of twenty 11-day embryos expressed beta-galactosidase specifically, two of them coming from the same foster mother. In whole mount preparations all three were stained in the telencephalon and diencephalon, the mandibles, nostrils, midbrain and hindbrain (figures 5-33, 5-37, 5-40). Two of them exhibited reporter gene expression also in the heart and somites whereas one of these two also showed staining in the second brachial arch, the Wolfian duct and very faint staining in the left and right forelimbs. All the mice with beta-Gal staining had integrated the injected fragment in their genome as proved with

PCR analysis, however not all mice that had integrated this fragment in their genome showed also the staining. The variation in expression between the mice of the same developmental stage can be attributed to the fact that the injected construct is inserted randomly in the mouse genome. By being positioned in a silent and tightly packed chromatin region, it may show less or no expression or, in contrary, it can show expression in many tissues if inserted in a very active genomic site.

6.7 The reporter gene expression pattern appears to recapitulate GLI3 activity

From the beta-galactosidase staining results it can be deduced that human Hom2 has the potential to activate at developmental days 9 to 11 in mouse embryos reporter gene expression mostly in the brain, mandibles, nostrils and heart. *Gli3* plays an important role regulating the development of all of these organs and tissues (Hardcastle et al. 1998; Balmer and LaMantia 2004). Detailed expression patterns of Gli3 in 9 and 11 days mouse embryos are not readily available from publications. Gli3 antibodies suited for histochemical staining are apparently missing. Therefore, Gli3 expression had preferentially been studied by in situ hybridisation which detects the mRNA but does not prove that the translation product is present (Buscher and Ruether 1998; Aoto et al. 2002; Zoltewicz et al. 2004). In addition, the interest in Gli3 expression studies has been focused on limb development.

The expression patterns of the transgene in the mouse embryos carrying the Hom2 construct were ascertained with the very sensitive conversion of a chromogenic substrate by beta-galactosidase. In addition to the study of the whole mount aspect, histological sections were studied which allow assigning the gene expression to individual cells or groups of cells (figure 6-9, B). Comparably detailed data are not available for Gli3 expression during normal development. Therefore, a firm statement about the reporter gene expression recapitulating Gli3 expression can only be made concerning a coarse pattern, such as in figure 6-9 A and C, which demonstrate similarity between Gli3 expression in the telencephalon and the pattern of transgene expression.

Evolutionarily conserved enhancer elements recapitulating in transgenic constructs with reporter genes the expression of the gene that they normally regulate have recently been reported for other developmental genes such as the *DACH* gene.

Dissecting these elements allows to assign the regulatory potential for expression in individual tissues and at given timepoints very specifically to small subregions possibly associated with appropriate transcription factors (Nobrega et al. 2003).

The results obtained in the present study with the Hom2 region need to be extended to more transgenic embryos and to be studied throughout all embryonic stages to uncover the complete regulatory potential of the Hom2 enhancer.

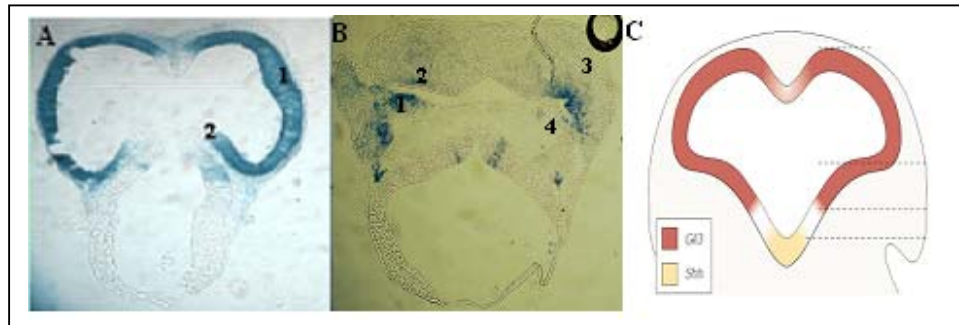


Figure 6-9: Head sections of day 10 and 11 mouse embryos comparing Hom2-lacZ activity with the pattern of Gli3 and Shh expression

Horizontal head sections of an 11 day, Hom2-lacZ expressing mouse. **A1**: Wall of telencephalic vesicle. **A2**: Wall of diencephalon. **B1**: Trigeminal (V) neural crest cells migrating into maxillary component of first branchial arch. **B2**: Maxillary component of first branchial arch. **B3**: Mandibular component of first branchial arch. **B4**: Trigeminal (V) ganglion. **C**: Expression domains of *GLI3* (red) and *Shh* (yellow) in the telencephalon; diagram of a horizontal head section of a 10 day embryo (Rallu et al. 2002). The expression of *GLI3* resembles the beta-galactosidase staining in A.

In addition to the study of the whole mount aspect, histological sections were studied which allow assigning the gene expression to individual cells or groups of cells (figure 6-9, B).

Taking a look at the results obtained with transient transfection into cell cultures allows to speculate about potential transcription factors which might be associated with Hom2 activity in vivo: Strong expression of NFATp, the regulatory candidate for Hom2 studied in cell culture, had been detected with northern blots in leukocytes, spleen, and placenta, intermediate expression in thymus, prostate, testis, ovary, small intestine, lung, liver, muscle, kidney, and pancreas, and very weak expression in heart, colon, and brain (Masuda et al. 1995). Thus, the factor is present in some tissues in which Hom2 drove expression of lacZ. However, NFATc expression had neither been

detected in the brain nor had NFATx been demonstrated in both the brain and the heart (Masuda et al. 1995). These two factors were also predicted to bind at the sites that were mutated in the sequence of Hom2.

Other proteins are also predicted to bind to the mutated sites in Hom2. For instance, c-Ets-2 is highly expressed in the newly forming cartilage, in the skull precursor cells and vertebral primordial. Overexpression of this gene results in skeletal and lymphocyte abnormalities and in a smaller thymus (Sumarsono et al. 1996; Wolvetang et al. 2003). NF1 participates in bone formation and has a general role in neural crest-derived tissues (Cichowski and Jacks 2001). It remains to be seen if Hom2 directs reporter gene expression in these tissues at other times during development, which would add credibility to the putative involvement of these proteins in *GLI3* regulation.

In some cases, a cis-acting regulator element may reside at a long distance upstream or downstream of its target, even inside an intron of a neighbouring gene (Gilligan and Venkatesh 2004). Therefore, the genomic segment around *GLI3* needs to be scrutinized for potential other regulatory targets of Hom2. *GLI3* is located within a “gene desert, i.e. few genes are observed in close neighbourhood. Within a segment of 1Mb telomeric of *GLI3* only the genes of two hypothetical proteins and of the inhibin beta A (INHBA) are located. The inhibin beta A subunit joins the alpha subunit to form a pituitary FSH secretion inhibitor, and its mutation affects craniofacial features but not the development or function of the brain and the heart (Matzuk et al. 1995), regions in which Hom2 shows transcriptional regulatory activity. 1 Mb centromeric of *GLI3* two more genes of hypothetical proteins, the mitochondrial ribosomal protein L32 (MRPL32) and the proteasome alpha 2 subunit (PSMA2) can be found. The latter ones are expressed ubiquitously and, therefore, it is highly unlikely that Hom2 is involved in their expression.

It cannot be excluded that Hom2 is a regulator for an even more distantly located gene or that it is an element that acts positively on more than one gene, including *GLI3*. Translocation events at distant sites on a chromosome associated with a specific clinical phenotype, such as described for preaxial polydactyly (PPD) (Lettice et al. 2002) hint at distant regulatory elements of a gene. However, no clinical phenotypes other than *GLI3* morphopathies have been associated with translocations affecting chromosome 7p within or near *GLI3*. Therefore, an effect of an element within *GLI3*

upon the neighbouring genes, which are expressed in patterns very different from the GLI3 specific pattern, is unlikely.

6.8 Hom2 is a conserved non-genic sequence element acting as transcriptional enhancer

The identification of cis-acting regulatory elements, traditionally, used to involve the systematic screening of large cloned segments, such as BACs or even YACs for their potential to activate gene expression in transient transfection studies with cell cultures or after transfer into animals such as mice, rabbits, pigs and other mammals (Giraldo and Montoliu 2001; Ristevski 2005). Segments showing this property had to be subdivided until the relevant sequences could be cloned and confirmed by mutagenesis (Niemann and Kues 2000).

Since the genomic sequence of many species is becoming available, comparative genomics showed that regulatory elements are enriched among highly conserved non-genic sequence elements. For instance, a report of an injected YAC construct carrying a conserved region, three interleukin genes and a non interleukin gene showed regulatory control of the conserved region specific for the interleukin genes (Hardison 2000). Unexpectedly, mammalian genomes contain many highly conserved sequences that are not functionally transcribed. They are single copy and comprise 1-2% of the human genome. These conserved sequences, which must be over 100 bp long and must have a similarity greater than 70% (between human and mouse) in order to be classified as CNGs (Conserved Non-Genic sequences), are found in humans, mice, dogs and most probably in the majority of mammals (Dermitzakis et al. 2002; Dermitzakis et al. 2005). These parameters comprise an arbitrary criterion of conservation that is frequently used as a reasonable empirical significance threshold in many biological studies. Characteristics of these elements are that the level of conservation and the clustering of substitutions are not correlated with distance from genes, with location in gene-rich versus gene-poor regions or with location in introns or between genes. In CNGs there is a tendency of AT to change into GC. In addition there is a higher incidence of CNGs in gene-poor regions, and they are more conserved than coding regions. (Dermitzakis et al. 2004; Dermitzakis et al. 2005).

The homologies studied in this thesis satisfy the conservation criteria to be characterized as CNGs. Especially, the sequence of Hom2 had already been identified

as an ultra-conserved fragment of chromosome 7 in a previous study in which 481 fragments of all the human genome were identified with 100% similarity between human, mouse and rat for a length greater than 200bp (Bejerano et al. 2004). The total sequence of Hom2 (434 bp) aligns un-gapped with the mouse sequence. Hom3 and Hom4 sequences have small gaps of inconsistency between human and mouse. Therefore they were not predicted under such stringent conditions used by Bejerano and colleagues (2004). These authors reasoned that if independent substitutions happened at each site, with the slowest neutral substitution rate that is observed for any 1 Mb region of the genome, then the chance to find only one of such ultra-conserved elements in the whole genome would be less than 10^{-22} . This number strongly suggests that the existence of such elements is not a phenomenon of chance. However, what might be the role for such conserved elements, some of which are conserved even down to the genome of fishes, remains an open question that may need to be answered for each element individually.

There are an increasing number of reports on such conserved elements that can regulate gene expression. *DACH* is a gene bracketed by two gene deserts 870 kb and 1330 kb in length. Its complex expression in numerous tissues speaks for strict regulation by distinct regulatory elements. Comparison of the genomic region encompassing this gene in man, mouse, frog, zebrafish and two pufferfish species identified 32 elements in a 2.6 Mb genomic region containing the *DACH* gene. Seven out of nine tested elements showed staining in transgenic mice in different tissues where *DACH* is also expressed (Nobrega et al. 2003).

Comparative analysis of the *SIM2* gene interval on human chromosome 21 with horse, cow, pig, dog, cat and mouse DNA resulted in a number of conserved elements which could up-regulate the *SIM2* promoter luciferase reporter construct in transiently transfected cell assays (Frazer et al. 2004). Hom2 does not seem to up-regulate the *GLI3* minimal promoter in similar assays but it can function by itself as a promoter. However, Hom3 and Hom4 appear to have a negative effect on the activity of this promoter.

The CNGs can also function as structural components of chromosomes. 11% of the conserved noncoding DNA consists of predicted MARs (Matrix-scaffold Attachment Regions) (Glazko et al. 2003). MARs are involved in the attachment of the chromatin to the nuclear matrix, chromatin remodelling and transcription regulation. In addition, more than half of the predicted MARs co-occur with one or more independently

identified conserved sequence blocks. Most of these predicted conserved MARs are found in regions 5' upstream of genes (Glazko et al. 2003). None of the three tested homologies in this work contain any MARs specific sequence motifs.

Still, the role of most of these highly conserved elements is unclear. As discussed above, one potential role of them is the regulation of gene expression, but they are considered to be too long and too conserved to contain densely packed transcription factor binding sites, which are generally degenerate. In addition, the independence from their genic environment makes it less likely that the majority of the CNGs are cis-regulatory elements. They might also be structural components of chromosomes but the fact that they are so highly conserved and are single-copy sequences makes it unlikely that their main role is the structure of chromatin. CNGs may regulate genes also in trans, as it is becoming more evident that chromosomes have specific locations in the nucleus during the cell cycle in mammalian cells. (Johnston and Stormo 2003; Dermitzakis et al. 2004). Although these reports classify conserved elements either as regulatory sequences or as structural components there is still a great number of such highly conserved regions especially between mammals which must be inspected for their role in life.

Of the three highly conserved elements studied here, Hom2 appears to be an autonomous regulatory sequence that can drive expression of the luciferase gene by itself, but can also activate the beta-globin promoter to regulate tissue and temporal specific lacZ expression in transgenic mice. For the other two conserved elements, Hom3 and Hom4, evidence for a negative regulatory potential on the GLI3 minimal promoter could only be obtained by transient expression of reporter constructs in cell cultures. It remains questionable to what extent cell cultures represent the normal situation in animals or humans. It will be difficult to establish transgenic animals that allow the analysis of negative regulatory effects. Since the criterion to categorize a CNG as a regulatory sequence is mainly detection of an activating potential, sequence elements binding in a given test system repressing factors might be missed.

Even the expression pattern of a transgene observed in animals might not reflect reliably the expression pattern of the original target gene. A conserved stem cell leukaemia haematopoietic (*SCL*) transcription factor enhancer has been reported of driving reporter gene expression in *SCL* negative tissues in addition to *SCL* expressing tissues showing that the genomic context of transgene insertion is also important for the correct expression of the gene (Gottgens et al. 2000).

Which species should be compared in the homology search to identify cis-acting regulatory elements for *Gli3*? In this project elements were analysed that are highly conserved between humans, mice and the pufferfish. It can be predicted, that regulatory elements detected by this approach might regulate the expression in organs or tissues present in fishes and mammals. Sites that were added during the evolution of tetrapods, such as hands with fingers might go undetected. It will be interesting to see, if the very faint fore limb expression of the transgene under the control of Hom2 observed at day 11, at later stages will turn into the characteristic expression pattern of GLI3 during limb development.

As reviewed by (Boffelli et al. 2004), homologies obtained by comparison of sequences of very distant species (human-fugu) are mostly functional and are good candidates of regulating genes which play a pivotal role during the embryonic development. Many developmental processes are conserved across the animal kingdom, therefore the regulation of the participating genes might have been kept unchanged.

On the other hand, highly conserved sequences like the ones investigated in this work are only observed in a subset of the genes. Comparison of distantly related species inhibits prediction for elements that are specific for more closely related species. Elements governing tetrapod or even primate specific morphology might be hidden in sequences conserved between closely related species but missing in distant relatives. Comparison of more closely related organisms, such as human and mouse, has the problem that non-uniform evolution rates across the human genome lead to a high degree of conserved sequences between human and mouse. Thus it is difficult to predict which sequences are functional. Comparison of highly related species, such as human and other primates, increases dramatically the number of conserved regions. However, using phylogenetic shadowing between primates one can identify sites that evolve slower and which might harbour regulatory sequences that may be primate specific. Phylogenetic shadowing analyses multiple aligned sequences to identify sites that evolve slower. Each position in the multiple alignment is fitted to a phylogenetic model to calculate the likelihood that the sequence accumulates variations at a fast or slow rate. Positions with several differences in different branches of the phylogenetic tree are more likely to evolve faster, thus most probably they are less probable of containing functional sequences. Even sequences of only four different primates can be chosen for alignment as long as they are only distantly related (Boffelli et al.

2004). This method was already used to find regulatory elements for human lipoprotein A (*LPA*), which is one of a small set of genes that arose recently in the primate lineage and consequently found in only a subset of primates (Lawn et al. 1995; Boffelli et al. 2004). It might also be used to identify primate specific *GLI3* regulatory elements. Through shadowing it is possible to use even the sequence polymorphisms in human populations to annotate the human genome. Through the efforts to sequence several different human genomes this may become feasible, and regulating elements specific for humans may finally be identified.

7 OUTLOOK

The study of the transcriptional regulation of *GLI3* in this thesis has provided important insight into the complex regulatory mechanism of this key transcription factor of the hedgehog signalling pathway.

At the same time, it has opened up the avenue to a plethora of urgent questions. In this work, two initiator sites have been identified for *GLI3*, and the activity of its minimal promoter as well as the regulatory potential of highly conserved intronic regions have been tested in different cell lines. Potent transcription factors that regulate the promoter still have to be identified as site directed mutagenesis of 5 different regions, individually, did not strongly affect the activity of the promoter. However, when all of those sites are deleted the activity decreases. A combination of mutations at these 5 sites may lead to a better understanding of the *GLI3* promoter regulation. In addition, more tests with the mutated *inr* site as well as mutation of the second *inr* site can show if one of them is used preferably for transcription in a given test system.

Regarding the highly conserved regulatory Hom2 region, the impact of the transcription factor NFATp still has to be confirmed. By co-transfecting Hom2 with a construct that expresses NFATp it can be tested if the activity of this region is proportional to the dosage of this protein. In addition, a stable transgenic line of mice expressing lacZ under the control of Hom2 must be obtained to investigate throughout mouse development in which tissues and cells this region activates transcription. The orientation-dependent activity of Hom2, in combination with the existence of an *inr* site at the 3' end of its sequence gives this region promoter-like attributes, although no other typical promoter features could be found in its sequence. 3' RACE PCR starting from this *inr* site may help to find out if a hidden gene is located downstream of this sequence. In addition, testing the activity of the construct containing Hom2 upstream of the *GLI3* minimal promoter in different cell lines can help in understanding the function of this ultra-conserved region.

Site directed mutagenesis in the sequence of Hom3 and Hom4 will help to understand which factors bind to them and cause the repressive activity of these homologies. The sequence residing in the 1st intron, downstream of the promoter adjacent to the 1st exon is a further good candidate for a cis-acting regulatory element.

Sequencing these regions in the patients with GLI3 morphopathies, which do not show any mutation in their transcribed exons, will determine if mutations in those regulatory elements are associated with the clinical phenotype of GLI3 morphopathy. Finally, the 5' upstream region of *GLI2* still has to be characterized in greater detail and searched for a promoter and a transcription initiator site.

The conserved non-genic regions with regulatory potential identified within *GLI3* appear not to have paralogs anywhere inside, upstream or downstream of *GLI1* and *GLI2*. Thus, these paralog genes have evolved regulatory sites of their own, although they are expressed a coordinated fashion with each other. Determining if the regulatory elements of *GLI2* and *GLI1*, each on its own, share ancient homologies with orthologs in evolutionary distant species will cast light on the mechanism of speciation and the time point when the functional divergence of these genes evolved.

8 ABBREVIATIONS

a/a	aminoacid
AgNO ₃	Silver nitrate
APS	Ammonium peroxodisulfate
bp	base pair
BSA	Bovine Serum Albumin
cDNA	complementary DNA
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotidetriphosphate
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
FCS	Fetal Calf Serum
F	forward
g	gram
HCl	Hydrochloric acid
HEPES	1-Piperazineethane sulfonic acid
HPLC	High Performance Liquid Chromatography
IPTG	Isopropyl-β-thiogalactopyranoside
kb	kilo bases
kD	kilodalton
M	Molar
MgCl ₂	Magnesium chloride
mRNA	messenger Ribonucleic acid
NaCl	Sodium chloride
Na ₂ CO ₃	Sodium carbonate
ng	nanogram
nt	nucleotide
OMIM	Online Mendelian Inheritance in Man
PBS	Phosphate Buffer Saline
PCR	Polymerase Chain Reaction
pmol	picomol
R	reverse

RNA	Ribonucleic acid
RNase	Ribonuclease
rpm	Revolutions Per Minute
RT	Room Temperature
RT-PCR	Reverse Transcription-PCR
SDS	Sodium Dodecyl Sulfate
SSCA	Single Stranded Conformational Analysis
SV 40	Simian Virus 40
<i>Taq</i>	Polymerase isolated from <i>Thermus aquaticus</i>
TAE	Tris Acetate EDTA
TE	Tris EDTA
TEMED	Tetramethylethylene diamine
Tris	Tri-hydroxymethyl aminomethane
u	unit
UV	Ultra Violet
V	Voltage
ver	version
v/v	volume/volume
w/v	weight/volume

9 REFERENCES

- Alexandre C, Jacinto A, Ingham PW (1996) Transcriptional activation of hedgehog target genes in *Drosophila* is mediated directly by the cubitus interruptus protein, a member of the GLI family of zinc finger DNA-binding proteins. *Genes Dev* 10:2003-2013
- Arheden K, Ronne M, Mandahl N, Heim S, Kinzler KW, Vogelstein B, Mitelman F (1989) In situ hybridization localizes the human putative oncogene GLI to chromosome subbands 12q13.3-14.1. *Hum Genet* 82:1-2
- Aza-Blanc P, Kornberg TB (1999) Ci: a complex transducer of the hedgehog signal. *Trends Genet* 15:458-462
- Aza-Blanc P, Ramirez-Weber FA, Laget MP, Schwartz C, Kornberg TB (1997) Proteolysis that is inhibited by hedgehog targets Cubitus interruptus protein to the nucleus and converts it to a repressor. *Cell* 89:1043-1053
- Bai CB, Stephen D, Joyner AL (2004) All mouse ventral spinal cord patterning by hedgehog is Gli dependent and involves an activator function of Gli3. *Dev Cell* 6:103-115
- Bitgood MJ, McMahon AP (1995) Hedgehog and Bmp genes are coexpressed at many diverse sites of cell-cell interaction in the mouse embryo. *Dev Biol* 172:126-138
- Bitgood MJ, Shen L, McMahon AP (1996) Sertoli cell signaling by Desert hedgehog regulates the male germline. *Curr Biol* 6:298-304
- Brewster R, Mullor JL, Ruiz i Altaba A (2000) Gli2 functions in FGF signaling during antero-posterior patterning. *Development* 127:4395-4405
- Buckingham M, Alonso S, Barton P, Cohen A, Daubas P, Garner I, Robert B, Weydert A (1986) Actin and myosin multigene families: their expression during the formation and maturation of striated muscle. *Am J Med Genet* 25:623-634
- Bulgakov OV, Eggenschwiler JT, Hong DH, Anderson KV, Li T (2004) FKBP8 is a negative regulator of mouse sonic hedgehog signaling in neural tissues. *Development* 131:2149-2159
- Buescher D, Ruether U (1998) Expression profile of Gli family members and Shh in normal and mutant mouse limb development. *Dev Dyn* 211:88-96
- Callahan CA, Ofstad T, Horng L, Wang JK, Zhen HH, Coulombe PA, Oro AE (2004) MIM/BEG4, a Sonic hedgehog-responsive gene that potentiates Gli-dependent transcription. *Genes Dev* 18:2724-2729
- Chen Y, Gallaher N, Goodman RH, Smolik SM (1998) Protein kinase A directly regulates the activity and proteolysis of cubitus interruptus. *Proc Natl Acad Sci U S A* 95:2349-2354
- Cheng SY, Bishop JM (2002) Suppressor of Fused represses Gli-mediated transcription by recruiting the SAP18-mSin3 corepressor complex. *Proc Natl Acad Sci U S A* 99:5442-5447
- Chuang PT, Kornberg TB (2000) On the range of hedgehog signaling. *Curr Opin Genet Dev* 10:515-522
- Chuang PT, McMahon AP (1999) Vertebrate Hedgehog signalling modulated by induction of a Hedgehog-binding protein. *Nature* 397:617-621
- Cora D, Herrmann C, Dieterich C, Di Cunto F, Provero P, Caselle M (2005) Ab initio identification of putative human transcription factor binding sites by comparative genomics. *BMC Bioinformatics* 6:110

- Dahmane N, Lee J, Robins P, Heller P, Ruiz i Altaba A (1997) Activation of the transcription factor Gli1 and the Sonic hedgehog signalling pathway in skin tumours. *Nature* 389:876-881
- Dai P, Akimaru H, Tanaka Y, Maekawa T, Nakafuku M, Ishii S (1999) Sonic Hedgehog-induced activation of the Gli1 promoter is mediated by GLI3. *J Biol Chem* 274:8143-8152
- Denef N, Neubuser D, Perez L, Cohen SM (2000) Hedgehog induces opposite changes in turnover and subcellular localization of patched and smoothened. *Cell* 102:521-531
- Donner AL, Maas RL (2004) Conservation and non-conservation of genetic pathways in eye specification. *Int J Dev Biol* 48:743-753
- Eggenchwil JT, Espinoza E, Anderson KV (2001) Rab23 is an essential negative regulator of the mouse Sonic hedgehog signalling pathway. *Nature* 412:194-198
- Fan CM, Porter JA, Chiang C, Chang DT, Beachy PA, Tessier-Lavigne M (1995) Long-range sclerotome induction by sonic hedgehog: direct role of the amino-terminal cleavage product and modulation by the cyclic AMP signaling pathway. *Cell* 81:457-465
- Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151:1531-1545
- Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I (2004) VISTA: computational tools for comparative genomics. *Nucleic Acids Res* 32:W273-279
- Ghali L, Wong ST, Green J, Tidman N, Quinn AG (1999) Gli1 protein is expressed in basal cell carcinomas, outer root sheath keratinocytes and a subpopulation of mesenchymal cells in normal human skin. *J Invest Dermatol* 113:595-599
- Giles RH, Peters DJ, Breuning MH (1998) Conjunction dysfunction: CBP/p300 in human disease. *Trends Genet* 14:178-183
- Glise B, Miller CA, Crozatier M, Halbisen MA, Wise S, Olson DJ, Vincent A, Blair SS (2005) Shifted, the Drosophila ortholog of Wnt inhibitory factor-1, controls the distribution and movement of Hedgehog. *Dev Cell* 8:255-266
- Gomez-Skarmeta JL, Rodriguez I, Martinez C, Culi J, Ferres-Marco D, Beamonte D, Modolell J (1995) Cis-regulation of achaete and scute: shared enhancer-like elements drive their coexpression in proneural clusters of the imaginal discs. *Genes Dev* 9:1869-1882
- Gorfinkiel N, Sierra J, Callejo A, Ibanez C, Guerrero I (2005) The Drosophila ortholog of the human Wnt inhibitor factor Shifted controls the diffusion of lipid-modified Hedgehog. *Dev Cell* 8:241-253
- Grachtchouk M, Mo R, Yu S, Zhang X, Sasaki H, Hui CC, Dlugosz AA (2000) Basal cell carcinomas in mice overexpressing Gli2 in skin. *Nat Genet* 24:216-217
- Han C, Belenkaya TY, Wang B, Lin X (2004) Drosophila glypicans control the cell-to-cell movement of Hedgehog by a dynamin-independent process. *Development* 131:601-611
- Hardcastle Z, Mo R, Hui CC, Sharpe PT (1998) The Shh signalling pathway in tooth development: defects in Gli2 and Gli3 mutants. *Development* 125:2803-2811
- Hooper JE, Scott MP (2005) Communicating with Hedgehogs. *Nat Rev Mol Cell Biol* 6:306-317

- Hsieh JC, Kodjabachian L, Rebbert ML, Rattner A, Smallwood PM, Samos CH, Nusse R, Dawid IB, Nathans J (1999) A new secreted protein that binds to Wnt proteins and inhibits their activities. *Nature* 398:431-436
- Huangfu D, Liu A, Rakeman AS, Murcia NS, Niswander L, Anderson KV (2003) Hedgehog signalling in the mouse requires intraflagellar transport proteins. *Nature* 426:83-87
- Hughes DC, Allen J, Morley G, Sutherland K, Ahmed W, Prosser J, Lettice L, Allan G, Mattei MG, Farrall M, Hill RE (1997) Cloning and sequencing of the mouse Gli2 gene: localization to the Dominant hemimelia critical region. *Genomics* 39:205-215
- Inaki M, Kojima T, Ueda R, Saigo K (2002) Requirements of high levels of Hedgehog signaling activity for medial-region cell fate determination in *Drosophila* legs: identification of pxb, a putative Hedgehog signaling attenuator gene repressed along the anterior-posterior compartment boundary. *Mech Dev* 116:3-18
- Incardona JP, Lee JH, Robertson CP, Enga K, Kapur RP, Roelink H (2000) Receptor-mediated endocytosis of soluble and membrane-tethered Sonic hedgehog by Patched-1. *Proc Natl Acad Sci U S A* 97:12044-12049
- Ingham PW (1998) Transducing Hedgehog: the story so far. *Embo J* 17:3505-3511
- Ishibashi M, Saitsu H, Komada M, Shiota K (2005) Signaling cascade coordinating growth of dorsal and ventral tissues of the vertebrate brain, with special reference to the involvement of Sonic Hedgehog signaling. *Anat Sci Int* 80:30-36
- Ishikawa N, Shimada N, Takagi Y, Ishijima Y, Fukuda M, Kimura N (2003) Molecular evolution of nucleoside diphosphate kinase genes: conserved core structures and multiple-layered regulatory regions. *J Bioenerg Biomembr* 35:7-18
- Iwamoto M, Enomoto-Iwamoto M, Kurisu K (1999) Actions of hedgehog proteins on skeletal cells. *Crit Rev Oral Biol Med* 10:477-486
- Izraeli S, Lowe LA, Bertness VL, Campaner S, Hahn H, Kirsch IR, Kuehn MR (2001) Genetic evidence that Sil is required for the Sonic Hedgehog response pathway. *Genesis* 31:72-77
- Jacob J, Briscoe J (2003) Gli proteins and the control of spinal-cord patterning. *EMBO Rep* 4:761-765
- Jacobsen E-M (1996) Charakterisierung einer Kontrollregion des Kandidatengens für das Greig-Syndrom. Diplomwork, Philipps-University Marburg, Germany
- Jiang J, Struhl G (1998) Regulation of the Hedgehog and Wingless signalling pathways by the F-box/WD40-repeat protein Slimb. *Nature* 391:493-496
- Johnston JJ, Olivos-Glander I, Killoran C, Elson E, Turner JT, Peters KF, Abbott MH, et al. (2005) Molecular and clinical analyses of Greig cephalopolysyndactyly and Pallister-Hall syndromes: robust phenotype prediction from the type and position of GLI3 mutations. *Am J Hum Genet* 76:609-622
- Kalderon D (2000) Transducing the hedgehog signal. *Cell* 103:371-374
- Kalff-Suske M, Wild A, Topp J, Wessling M, Jacobsen EM, Bornholdt D, Engel H, Heuer H, Aalfs CM, Ausems MG, Barone R, Herzog A, Heutink P, Homfray T, Gillessen-Kaesbach G, Koenig R, Kunze J, Meinecke P, Muller D, Rizzo R, Streng S, Superti-Furga A, Grzeschik KH (1999) Point mutations throughout the GLI3 gene cause Greig cephalopolysyndactyly syndrome. *Hum Mol Genet* 8:1769-1777

- Kang S, Graham JM, Jr., Olney AH, Biesecker LG (1997) GLI3 frameshift mutations cause autosomal dominant Pallister-Hall syndrome. *Nat Genet* 15:266-268
- Karlstrom RO, Tyurina OV, Kawakami A, Nishioka N, Talbot WS, Sasaki H, Schier AF (2003) Genetic analysis of zebrafish *gli1* and *gli2* reveals divergent requirements for gli genes in vertebrate development. *Development* 130:1549-1564
- Karpen HE, Bukowski JT, Hughes T, Gratton JP, Sessa WC, Gailani MR (2001) The sonic hedgehog receptor patched associates with caveolin-1 in cholesterol-rich microdomains of the plasma membrane. *J Biol Chem* 276:19503-19511
- Kas K, Wlodarska I, Meyen E, Van den Berghe H, Van de Ven WJ (1996) Assignment of the gene encoding human Kruppel-related zinc finger protein 4 (GLI4) to 8q24.3 by fluorescent in situ hybridization. *Cytogenet Cell Genet* 72:297-298
- Katoh Y, Katoh M (2004) KIF27 is one of orthologs for Drosophila Costal-2. *Int J Oncol* 25:1875-1880
- Kinzler KW, Bigner SH, Bigner DD, Trent JM, Law ML, O'Brien SJ, Wong AJ, Vogelstein B (1987) Identification of an amplified, highly expressed gene in a human glioma. *Science* 236:70-73
- Kinzler KW, Ruppert JM, Bigner SH, Vogelstein B (1988) The GLI gene is a member of the Kruppel family of zinc finger proteins. *Nature* 332:371-374
- Kinzler KW, Vogelstein B (1990) The GLI gene encodes a nuclear protein which binds specific sequences in the human genome. *Mol Cell Biol* 10:634-642
- Kleinjan DA, van Heyningen V (2005) Long-range control of gene expression: emerging mechanisms and disruption in disease. *Am J Hum Genet* 76:8-32
- Kogerman P, Grimm T, Kogerman L, Krause D, Unden AB, Sandstedt B, Toftgard R, Zaphiropoulos PG (1999) Mammalian suppressor-of-fused modulates nuclear-cytoplasmic shuttling of Gli-1. *Nat Cell Biol* 1:312-319
- Krauss S, Concordet JP, Ingham PW (1993) A functionally conserved homolog of the Drosophila segment polarity gene *hh* is expressed in tissues with polarizing activity in zebrafish embryos. *Cell* 75:1431-1444
- Kuijper S, Beverdam A, Kroon C, Brouwer A, Candille S, Barsh G, Meijlink F (2005) Genetics of shoulder girdle formation: roles of *Tbx15* and *aristaless*-like genes. *Development* 132:1601-1610
- Lettice LA, Heaney SJ, Purdie LA, Li L, de Beer P, Oostra BA, Goode D, Elgar G, Hill RE, de Graaff E (2003) A long-range *Shh* enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet* 12:1725-1735
- Litingtung Y, Dahn RD, Li Y, Fallon JF, Chiang C (2002) *Shh* and *Gli3* are dispensable for limb skeleton formation but regulate digit number and identity. *Nature* 418:979-983
- Liu CZ, Yang JT, Yoon JW, Villavicencio E, Pfendler K, Walterhouse D, Iannaccone P (1998) Characterization of the promoter region and genomic organization of *GLI*, a member of the Sonic hedgehog-Patched signaling pathway. *Gene* 209:1-11
- Liu T, Wu J, He F (2000) Evolution of cis-acting elements in 5' flanking regions of vertebrate actin genes. *J Mol Evol* 50:22-30
- Manzanares M, Cordes S, Ariza-McNaughton L, Sadl V, Maruthainar K, Barsh G, Krumlauf R (1999) Conserved and distinct roles of *kreisler* in regulation of the paralogous *Hoxa3* and *Hoxb3* genes. *Development* 126:759-769

- Marigo V, Johnson RL, Vortkamp A, Tabin CJ (1996) Sonic hedgehog differentially regulates expression of GLI and GLI3 during limb development. *Dev Biol* 180:273-283
- Marigo V, Roberts DJ, Lee SM, Tsukurov O, Levi T, Gastier JM, Epstein DJ, Gilbert DJ, Copeland NG, Seidman CE, et al. (1995) Cloning, expression, and chromosomal location of SHH and IHH: two human homologues of the *Drosophila* segment polarity gene hedgehog. *Genomics* 28:44-51
- Marti E, Bovolenta P (2002) Sonic hedgehog in CNS development: one signal, multiple outputs. *Trends Neurosci* 25:89-96
- Matsumoto N, Fujimoto M, Kato R, Niikawa N (1996) Assignment of the human GLI2 gene to 2q14 by fluorescence in situ hybridization. *Genomics* 36:220-221
- McCarthy RA, Barth JL, Chintalapudi MR, Knaak C, Argraves WS (2002) Megalin functions as an endocytic sonic hedgehog receptor. *J Biol Chem* 277:25660-25667
- McDermott A, Gustafsson M, Elsam T, Hui CC, Emerson CP, Jr., Borycki AG (2005) Gli2 and Gli3 have redundant and context-dependent function in skeletal muscle formation. *Development* 132:345-357
- McMahon AP (2000) More surprises in the Hedgehog signaling pathway. *Cell* 100:185-188
- Merchant M, Vajdos FF, Ultsch M, Maun HR, Wendt U, Cannon J, Desmarais W, Lazarus RA, de Vos AM, de Sauvage FJ (2004) Suppressor of fused regulates Gli activity through a dual binding mechanism. *Mol Cell Biol* 24:8627-8641
- Methot N, Basler K (2000) Suppressor of fused opposes hedgehog signal transduction by impeding nuclear accumulation of the activator form of Cubitus interruptus. *Development* 127:4001-4010
- Mo R, Freer AM, Zinyk DL, Crackower MA, Michaud J, Heng HH, Chik KW, Shi XM, Tsui LC, Cheng SH, Joyner AL, Hui C (1997) Specific and redundant functions of Gli2 and Gli3 zinc finger genes in skeletal patterning and development. *Development* 124:113-123
- Monnier V, Dussillol F, Alves G, Lamour-Isnard C, Plessis A (1998) Suppressor of fused links fused and Cubitus interruptus on the hedgehog signalling pathway. *Curr Biol* 8:583-586
- Morrison A, Ariza-McNaughton L, Gould A, Featherstone M, Krumlauf R (1997) HOXD4 and regulation of the group 4 paralog genes. *Development* 124:3135-3146
- Motoyama J, Liu J, Mo R, Ding Q, Post M, Hui CC (1998) Essential function of Gli2 and Gli3 in the formation of lung, trachea and oesophagus. *Nat Genet* 20:54-57
- Motoyama J, Milenkovic L, Iwama M, Shikata Y, Scott MP, Hui CC (2003) Differential requirement for Gli2 and Gli3 in ventral neural cell fate specification. *Dev Biol* 259:150-161
- Nobrega MA, Pennacchio LA (2004) Comparative genomic analysis as a tool for biological discovery. *J Physiol* 554:31-39
- Paces-Fessy M, Boucher D, Petit E, Paute-Briand S, Blanchet-Tournier MF (2004) The negative regulator of Gli, Suppressor of fused (Sufu), interacts with SAP18, Galectin3 and other nuclear proteins. *Biochem J* 378:353-362
- Park HL, Bai C, Platt KA, Matisse MP, Beeghly A, Hui CC, Nakashima M, Joyner AL (2000) Mouse Gli1 mutants are viable but have defects in SHH signaling in combination with a Gli2 mutation. *Development* 127:1593-1605

- Pavletich NP, Pabo CO (1993) Crystal structure of a five-finger GLI-DNA complex: new perspectives on zinc fingers. *Science* 261:1701-1707
- Pepinsky RB, Zeng C, Wen D, Rayhorn P, Baker DP, Williams KP, Bixler SA, Ambrose CM, Garber EA, Miatkowski K, Taylor FR, Wang EA, Galdes A (1998) Identification of a palmitic acid-modified form of human Sonic hedgehog. *J Biol Chem* 273:14037-14045
- Perrimon N (1995) Hedgehog and beyond. *Cell* 80:517-520
- Rahnama F, Toftgard R, Zaphiropoulos PG (2004) Distinct roles of PTCH2 splice variants in Hedgehog signalling. *Biochem J* 378:325-334
- Rallis C, Del Buono J, Logan MP (2005) Tbx3 can alter limb position along the rostrocaudal axis of the developing embryo. *Development* 132:1961-1970
- Riddle RD, Johnson RL, Laufer E, Tabin C (1993) Sonic hedgehog mediates the polarizing activity of the ZPA. *Cell* 75:1401-1416
- Robbins DJ, Nybakken KE, Kobayashi R, Sisson JC, Bishop JM, Therond PP (1997) Hedgehog elicits signal transduction by means of a large complex containing the kinesin-related protein costal2. *Cell* 90:225-234
- Roberts WM, Douglass EC, Peiper SC, Houghton PJ, Look AT (1989) Amplification of the gli gene in childhood sarcomas. *Cancer Res* 49:5407-5413
- Roelink H, Porter JA, Chiang C, Tanabe Y, Chang DT, Beachy PA, Jessell TM (1995) Floor plate and motor neuron induction by different concentrations of the amino-terminal cleavage product of sonic hedgehog autoproteolysis. *Cell* 81:445-455
- Roessler E, Du YZ, Mullor JL, Casas E, Allen WP, Gillessen-Kaesbach G, Roeder ER, Ming JE, Ruiz i Altaba A, Muenke M (2003) Loss-of-function mutations in the human GLI2 gene are associated with pituitary anomalies and holoprosencephaly-like features. *Proc Natl Acad Sci U S A* 100:13424-13429
- Ruiz i Altaba A (1999) Gli proteins encode context-dependent positive and negative functions: implications for development and disease. *Development* 126:3205-3216
- Ruppert JM, Kinzler KW, Wong AJ, Bigner SH, Kao FT, Law ML, Seunanez HN, O'Brien SJ, Vogelstein B (1988) The GLI-Kruppel family of human genes. *Mol Cell Biol* 8:3104-3113
- Ruppert JM, Vogelstein B, Arheden K, Kinzler KW (1990) GLI3 encodes a 190-kilodalton protein with multiple regions of GLI similarity. *Mol Cell Biol* 10:5408-5415
- Sagai T, Hosoya M, Mizushima Y, Tamura M, Shiroishi T (2005) Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. *Development* 132:797-803
- Salinas PC (2003) The morphogen sonic hedgehog collaborates with netrin-1 to guide axons in the spinal cord. *Trends Neurosci* 26:641-643
- Sasaki H, Nishizaki Y, Hui C, Nakafuku M, Kondoh H (1999) Regulation of Gli2 and Gli3 activities by an amino-terminal repression domain: implication of Gli2 and Gli3 as primary mediators of Shh signaling. *Development* 126:3915-3924
- Schwartz C, Locke J, Nishida C, Kornberg TB (1995) Analysis of cubitus interruptus regulation in *Drosophila* embryos and imaginal disks. *Development* 121:1625-1635
- Schwartz S, Zhang Z, Frazer KA, Smit A, Riemer C, Bouck J, Gibbs R, Hardison R, Miller W (2000) PipMaker--a web server for aligning two genomic DNA sequences. *Genome Res* 10:577-586

- Schweitzer R, Vogan KJ, Tabin CJ (2000) Similar expression and regulation of Gli2 and Gli3 in the chick limb bud. *Mech Dev* 98:171-174
- Sekimizu K, Nishioka N, Sasaki H, Takeda H, Karlstrom RO, Kawakami A (2004) The zebrafish iguana locus encodes Dzip1, a novel zinc-finger protein required for proper regulation of Hedgehog signaling. *Development* 131:2521-2532
- Sisson JC, Ho KS, Suyama K, Scott MP (1997) Costal2, a novel kinesin-related protein in the Hedgehog signaling pathway. *Cell* 90:235-245
- St-Jacques B, Hammerschmidt M, McMahon AP (1999) Indian hedgehog signaling regulates proliferation and differentiation of chondrocytes and is essential for bone formation. *Genes Dev* 13:2072-2086
- Tanimura A, Dan S, Yoshida M (1998) Cloning of novel isoforms of the human Gli2 oncogene and their activities to enhance tax-dependent transcription of the human T-cell leukemia virus type 1 genome. *J Virol* 72:3958-3964
- Tay SY, Ingham PW, Roy S (2005) A homologue of the Drosophila kinesin-like protein Costal2 regulates Hedgehog signal transduction in the vertebrate embryo. *Development* 132:625-634
- Tyurina OV, Guner B, Popova E, Feng J, Schier AF, Kohtz JD, Karlstrom RO (2005) Zebrafish Gli3 functions as both an activator and a repressor in Hedgehog signaling. *Dev Biol* 277:537-556
- Venkatesh B, Yap WH (2005) Comparative genomics using fugu: a tool for the identification of conserved vertebrate cis-regulatory elements. *Bioessays* 27:100-107
- Villavicencio EH, Walterhouse DO, Iannaccone PM (2000) The sonic hedgehog-patched-gli pathway in human development and disease. *Am J Hum Genet* 67:1047-1054
- Vortkamp A, Gessler M, Grzeschik KH (1991) GLI3 zinc-finger gene interrupted by translocations in Greig syndrome families. *Nature* 352:539-540
- Vortkamp A, Gessler M, Grzeschik KH (1995) Identification of optimized target sequences for the GLI3 zinc finger protein. *DNA Cell Biol* 14:629-634
- Vortkamp A, Heid C, Gessler M, Grzeschik KH (1995) Isolation and characterization of a cosmid contig for the GCPS gene region. *Hum Genet* 95:82-88
- Vortkamp A, Lee K, Lanske B, Segre GV, Kronenberg HM, Tabin CJ (1996) Regulation of rate of cartilage differentiation by Indian hedgehog and PTH-related protein. *Science* 273:613-622
- Wang B, Fallon JF, Beachy PA (2000) Hedgehog-regulated processing of Gli3 produces an anterior/posterior repressor gradient in the developing vertebrate limb. *Cell* 100:423-434
- Wang QT, Holmgren RA (1999) The subcellular localization and activity of Drosophila cubitus interruptus are regulated at multiple levels. *Development* 126:5097-5106
- Warburton D, Lee MK (1999) Current concepts on lung development. *Curr Opin Pediatr* 11:188-192
- Werner CA, Dohner H, Joos S, Trumper LH, Baudis M, Barth TF, Ott G, Moller P, Lichter P, Bentz M (1997) High-level DNA amplifications are common genetic aberrations in B-cell neoplasms. *Am J Pathol* 151:335-342
- Wild A, Kalff-Suske M, Vortkamp A, Bornholdt D, Koenig R, Grzeschik KH (1997) Point mutations in human GLI3 cause Greig syndrome. *Hum Mol Genet* 6:1979-1984
- Wolff C, Roy S, Lewis KE, Schauerte H, Joerg-Rauch G, Kirn A, Weiler C, Geisler R, Haffter P, Ingham PW (2004) iguana encodes a novel zinc-finger protein

- p with coiled-coil domains essential for Hedgehog signal transduction in the zebrafish embryo.
- Genes Dev*
- 18:1565-1576
- Yang JT, Liu CZ, Villavicencio EH, Yoon JW, Walterhouse D, Iannaccone PM (1997) Expression of human GLI in mice results in failure to thrive, early death, and patchy Hirschsprung-like gastrointestinal dilatation. *Mol Med* 3:826-835
- Alexandre C, Jacinto A, Ingham PW (1996) Transcriptional activation of hedgehog target genes in *Drosophila* is mediated directly by the cubitus interruptus protein, a member of the GLI family of zinc finger DNA-binding proteins. *Genes Dev* 10:2003-2013
- Aoto K, Nishimura T, Eto K, Motoyama J (2002) Mouse GLI3 regulates Fgf8 expression and apoptosis in the developing neural tube, face, and limb bud. *Dev Biol* 251:320-332
- Arheden K, Ronne M, Mandahl N, Heim S, Kinzler KW, Vogelstein B, Mitelman F (1989) In situ hybridization localizes the human putative oncogene GLI to chromosome subbands 12q13.3-14.1. *Hum Genet* 82:1-2
- Aza-Blanc P, Kornberg TB (1999) Ci: a complex transducer of the hedgehog signal. *Trends Genet* 15:458-462
- Aza-Blanc P, Ramirez-Weber FA, Laget MP, Schwartz C, Kornberg TB (1997) Proteolysis that is inhibited by hedgehog targets Cubitus interruptus protein to the nucleus and converts it to a repressor. *Cell* 89:1043-1053
- Azizkhan JC, Jensen DE, Pierce AJ, Wade M (1993) Transcription from TATA-less promoters: dihydrofolate reductase as a model. *Crit Rev Eukaryot Gene Expr* 3:229-254
- Bai CB, Stephen D, Joyner AL (2004) All mouse ventral spinal cord patterning by hedgehog is Gli dependent and involves an activator function of Gli3. *Dev Cell* 6:103-115
- Balmer CW, LaMantia AS (2004) Loss of Gli3 and Shh function disrupts olfactory axon trajectories. *J Comp Neurol* 472:292-307
- Bejerano G, Pheasant M, Makunin I, Stephen S, Kent WJ, Mattick JS, Haussler D (2004) Ultraconserved elements in the human genome. *Science* 304:1321-1325
- Bitgood MJ, McMahon AP (1995) Hedgehog and Bmp genes are coexpressed at many diverse sites of cell-cell interaction in the mouse embryo. *Dev Biol* 172:126-138
- Bitgood MJ, Shen L, McMahon AP (1996) Sertoli cell signaling by Desert hedgehog regulates the male germline. *Curr Biol* 6:298-304
- Boffelli D, Cheng JF, Rubin EM (2004) Convergent evolution in primates and an insectivore. *Genomics* 83:19-23
- Boffelli D, Nobrega MA, Rubin EM (2004) Comparative genomics at the vertebrate extremes. *Nat Rev Genet* 5:456-465
- Brewster R, Mullor JL, Ruiz i Altaba A (2000) Gli2 functions in FGF signaling during antero-posterior patterning. *Development* 127:4395-4405
- Brueton LA, Chotai KA, van Herwerden L, Schinzel A, Winter RM (1992) The acrocallosal syndrome and Greig syndrome are not allelic disorders. *J Med Genet* 29:635-637
- Buckingham M, Alonso S, Barton P, Cohen A, Daubas P, Garner I, Robert B, Weydert A (1986) Actin and myosin multigene families: their expression during the formation and maturation of striated muscle. *Am J Med Genet* 25:623-634

- Bulgakov OV, Eggenschwiler JT, Hong DH, Anderson KV, Li T (2004) FKBP8 is a negative regulator of mouse sonic hedgehog signaling in neural tissues. *Development* 131:2149-2159
- Buscher D, Ruether U (1998) Expression profile of Gli family members and Shh in normal and mutant mouse limb development. *Dev Dyn* 211:88-96
- Callahan CA, Ofstad T, Horng L, Wang JK, Zhen HH, Coulombe PA, Oro AE (2004) MIM/BEG4, a Sonic hedgehog-responsive gene that potentiates Gli-dependent transcription. *Genes Dev* 18:2724-2729
- Chen Y, Gallaher N, Goodman RH, Smolik SM (1998) Protein kinase A directly regulates the activity and proteolysis of cubitus interruptus. *Proc Natl Acad Sci U S A* 95:2349-2354
- Cheng SY, Bishop JM (2002) Suppressor of Fused represses Gli-mediated transcription by recruiting the SAP18-mSin3 corepressor complex. *Proc Natl Acad Sci U S A* 99:5442-5447
- Chuang PT, Kornberg TB (2000) On the range of hedgehog signaling. *Curr Opin Genet Dev* 10:515-522
- Chuang PT, McMahon AP (1999) Vertebrate Hedgehog signalling modulated by induction of a Hedgehog-binding protein. *Nature* 397:617-621
- Cichowski K, Jacks T (2001) NF1 tumor suppressor gene function: narrowing the GAP. *Cell* 104:593-604
- Cora D, Herrmann C, Dieterich C, Di Cunto F, Provero P, Caselle M (2005) Ab initio identification of putative human transcription factor binding sites by comparative genomics. *BMC Bioinformatics* 6:110
- Cowell IG, Skinner A, Hurst HC (1992) Transcriptional repression by a novel member of the bZIP family of transcription factors. *Mol Cell Biol* 12:3070-3077
- Dahmane N, Lee J, Robins P, Heller P, Ruiz i Altaba A (1997) Activation of the transcription factor Gli1 and the Sonic hedgehog signalling pathway in skin tumours. *Nature* 389:876-881
- Dai P, Akimaru H, Tanaka Y, Maekawa T, Nakafuku M, Ishii S (1999) Sonic Hedgehog-induced activation of the Gli1 promoter is mediated by GLI3. *J Biol Chem* 274:8143-8152
- Denef N, Neubuser D, Perez L, Cohen SM (2000) Hedgehog induces opposite changes in turnover and subcellular localization of patched and smoothened. *Cell* 102:521-531
- Dermitzakis ET, Kirkness E, Schwarz S, Birney E, Reymond A, Antonarakis SE (2004) Comparison of human chromosome 21 conserved nongenic sequences (CNGs) with the mouse and dog genomes shows that their selective constraint is independent of their genic environment. *Genome Res* 14:852-859
- Dermitzakis ET, Reymond A, Antonarakis SE (2005) Conserved non-genic sequences - an unexpected feature of mammalian genomes. *Nat Rev Genet* 6:151-157
- Dermitzakis ET, Reymond A, Lyle R, Scamuffa N, Ucla C, Deutsch S, Stevenson BJ, Flegel V, Bucher P, Jongeneel CV, Antonarakis SE (2002) Numerous potentially functional but non-genic conserved sequences on human chromosome 21. *Nature* 420:578-582
- Donner AL, Maas RL (2004) Conservation and non-conservation of genetic pathways in eye specification. *Int J Dev Biol* 48:743-753
- Durst KL, Hiebert SW (2004) Role of RUNX family members in transcriptional repression and gene silencing. *Oncogene* 23:4220-4224

- Eggenschwiler JT, Espinoza E, Anderson KV (2001) Rab23 is an essential negative regulator of the mouse Sonic hedgehog signalling pathway. *Nature* 412:194-198
- Eleouet JF, Romeo PH (1993) CCACC-binding or simian-virus-40-protein-1-binding proteins cooperate with human GATA-1 to direct erythroid-specific transcription and to mediate 5' hypersensitive site 2 sensitivity of a TATA-less promoter. *Eur J Biochem* 212:763-770
- Elson E, Perveen R, Donnai D, Wall S, Black GC (2002) De novo GLI3 mutation in acrocallosal syndrome: broadening the phenotypic spectrum of GLI3 defects and overlap with murine models. *J Med Genet* 39:804-806
- Emoto M, Miki M, Sarker AH, Nakamura T, Seki Y, Seki S, Ikeda S (2005) Structure and transcription promoter activity of mouse flap endonuclease 1 gene: Alternative splicing and bidirectional promoter. *Gene*
- Fabbro C, de Gemmis P, Braghetta P, Colombatti A, Volpin D, Bonaldo P, Bressan GM (2005) Analysis of regulatory regions of Emilin1 gene and their combinatorial contribution to tissue-specific transcription. *J Biol Chem* 280:15749-15760
- Fan CM, Porter JA, Chiang C, Chang DT, Beachy PA, Tessier-Lavigne M (1995) Long-range sclerotome induction by sonic hedgehog: direct role of the amino-terminal cleavage product and modulation by the cyclic AMP signaling pathway. *Cell* 81:457-465
- Fischer KD, Haese A, Nowock J (1993) Cooperation of GATA-1 and Sp1 can result in synergistic transcriptional activation or interference. *J Biol Chem* 268:23915-23923
- Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151:1531-1545
- Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I (2004) VISTA: computational tools for comparative genomics. *Nucleic Acids Res* 32:W273-279
- Frazer KA, Tao H, Osoegawa K, de Jong PJ, Chen X, Doherty MF, Cox DR (2004) Noncoding sequences conserved in a limited number of mammals in the SIM2 interval are frequently functional. *Genome Res* 14:367-372
- Ghali L, Wong ST, Green J, Tidman N, Quinn AG (1999) Gli1 protein is expressed in basal cell carcinomas, outer root sheath keratinocytes and a subpopulation of mesenchymal cells in normal human skin. *J Invest Dermatol* 113:595-599
- Giles RH, Peters DJ, Breuning MH (1998) Conjunction dysfunction: CBP/p300 in human disease. *Trends Genet* 14:178-183
- Gilligan P, Venkatesh B (2004) Application of comparative genomics to the analysis of vertebrate regulatory elements. *Brief Funct Genomic Proteomic* 3:7-11
- Giraldo P, Montoliu L (2001) Size matters: use of YACs, BACs and PACs in transgenic animals. *Transgenic Res* 10:83-103
- Glazko GV, Koonin EV, Rogozin IB, Shabalina SA (2003) A significant fraction of conserved noncoding DNA in human and mouse consists of predicted matrix attachment regions. *Trends Genet* 19:119-124
- Glise B, Miller CA, Crozatier M, Halbisen MA, Wise S, Olson DJ, Vincent A, Blair SS (2005) Shifted, the Drosophila ortholog of Wnt inhibitory factor-1, controls the distribution and movement of Hedgehog. *Dev Cell* 8:255-266
- Gomez-Skarmeta JL, Rodriguez I, Martinez C, Culi J, Ferres-Marco D, Beamonte D, Modolell J (1995) Cis-regulation of achaete and scute: shared enhancer-like

- elements drive their coexpression in proneural clusters of the imaginal discs. *Genes Dev* 9:1869-1882
- Gorfinkiel N, Sierra J, Callejo A, Ibanez C, Guerrero I (2005) The *Drosophila* ortholog of the human Wnt inhibitor factor Shifted controls the diffusion of lipid-modified Hedgehog. *Dev Cell* 8:241-253
- Gottgens B, Barton LM, Gilbert JG, Bench AJ, Sanchez MJ, Bahn S, Mistry S, Grafham D, McMurray A, Vaudin M, Amaya E, Bentley DR, Green AR, Sinclair AM (2000) Analysis of vertebrate SCL loci identifies conserved enhancers. *Nat Biotechnol* 18:181-186
- Grachtchouk M, Mo R, Yu S, Zhang X, Sasaki H, Hui CC, Dlugosz AA (2000) Basal cell carcinomas in mice overexpressing Gli2 in skin. *Nat Genet* 24:216-217
- Graef IA, Wang F, Charron F, Chen L, Neilson J, Tessier-Lavigne M, Crabtree GR (2003) Neurotrophins and netrins require calcineurin/NFAT signaling to stimulate outgrowth of embryonic axons. *Cell* 113:657-670
- Han C, Belenkaya TY, Wang B, Lin X (2004) *Drosophila* glypicans control the cell-to-cell movement of Hedgehog by a dynamin-independent process. *Development* 131:601-611
- Hardcastle Z, Mo R, Hui CC, Sharpe PT (1998) The Shh signalling pathway in tooth development: defects in Gli2 and Gli3 mutants. *Development* 125:2803-2811
- Hardison RC (2000) Conserved noncoding sequences are reliable guides to regulatory elements. *Trends Genet* 16:369-372
- Hooper JE, Scott MP (2005) Communicating with Hedgehogs. *Nat Rev Mol Cell Biol* 6:306-317
- Horsley V, Pavlath GK (2002) NFAT: ubiquitous regulator of cell differentiation and adaptation. *J Cell Biol* 156:771-774
- Hsieh JC, Kodjabachian L, Rebbert ML, Rattner A, Smallwood PM, Samos CH, Nusse R, Dawid IB, Nathans J (1999) A new secreted protein that binds to Wnt proteins and inhibits their activities. *Nature* 398:431-436
- Huangfu D, Liu A, Rakeman AS, Murcia NS, Niswander L, Anderson KV (2003) Hedgehog signalling in the mouse requires intraflagellar transport proteins. *Nature* 426:83-87
- Hughes DC, Allen J, Morley G, Sutherland K, Ahmed W, Prosser J, Lettice L, Allan G, Mattei MG, Farrall M, Hill RE (1997) Cloning and sequencing of the mouse Gli2 gene: localization to the Dominant hemimelia critical region. *Genomics* 39:205-215
- Hulme DJ, Blair IP, Dawkins JL, Nicholson GA (2000) Exclusion of NFIL3 as the gene causing hereditary sensory neuropathy type I by mutation analysis. *Hum Genet* 106:594-596
- Inaki M, Kojima T, Ueda R, Saigo K (2002) Requirements of high levels of Hedgehog signaling activity for medial-region cell fate determination in *Drosophila* legs: identification of pxb, a putative Hedgehog signaling attenuator gene repressed along the anterior-posterior compartment boundary. *Mech Dev* 116:3-18
- Incardona JP, Lee JH, Robertson CP, Enga K, Kapur RP, Roelink H (2000) Receptor-mediated endocytosis of soluble and membrane-tethered Sonic hedgehog by Patched-1. *Proc Natl Acad Sci U S A* 97:12044-12049
- Ingham PW (1998) Transducing Hedgehog: the story so far. *Embo J* 17:3505-3511
- Ishibashi M, Saito H, Komada M, Shiota K (2005) Signaling cascade coordinating growth of dorsal and ventral tissues of the vertebrate brain, with special

- reference to the involvement of Sonic Hedgehog signaling. *Anat Sci Int* 80:30-36
- Ishikawa N, Shimada N, Takagi Y, Ishijima Y, Fukuda M, Kimura N (2003) Molecular evolution of nucleoside diphosphate kinase genes: conserved core structures and multiple-layered regulatory regions. *J Bioenerg Biomembr* 35:7-18
- Iwamoto M, Enomoto-Iwamoto M, Kurisu K (1999) Actions of hedgehog proteins on skeletal cells. *Crit Rev Oral Biol Med* 10:477-486
- Izraeli S, Lowe LA, Bertness VL, Campaner S, Hahn H, Kirsch IR, Kuehn MR (2001) Genetic evidence that *Sil* is required for the Sonic Hedgehog response pathway. *Genesis* 31:72-77
- Jackson SM, Keech CA, Williamson DJ, Gutierrez-Hartmann A (1992) Interaction of basal positive and negative transcription elements controls repression of the proximal rat prolactin promoter in nonpituitary cells. *Mol Cell Biol* 12:2708-2719
- Jacob J, Briscoe J (2003) Gli proteins and the control of spinal-cord patterning. *EMBO Rep* 4:761-765
- Jacobsen E-M (1996) Charakterisierung einer Kontrollregion des Kandidatengens für das Greig-Syndrom. Diplomwork, Philipps-University Marburg, Germany
- Javahery R, Khachi A, Lo K, Zenzie-Gregory B, Smale ST (1994) DNA sequence requirements for transcriptional initiator activity in mammalian cells. *Mol Cell Biol* 14:116-127
- Jiang J, Struhl G (1998) Regulation of the Hedgehog and Wingless signalling pathways by the F-box/WD40-repeat protein Slimb. *Nature* 391:493-496
- Johnston JJ, Olivos-Glander I, Killoran C, Elson E, Turner JT, Peters KF, Abbott MH, et al. (2005) Molecular and clinical analyses of Greig cephalopolysyndactyly and Pallister-Hall syndromes: robust phenotype prediction from the type and position of *GLI3* mutations. *Am J Hum Genet* 76:609-622
- Johnston M, Stormo GD (2003) Evolution. Heirlooms in the attic. *Science* 302:997-999
- Kalderon D (2000) Transducing the hedgehog signal. *Cell* 103:371-374
- Kalff-Suske M, Wild A, Topp J, Wessling M, Jacobsen EM, Bornholdt D, Engel H, Heuer H, Aalfs CM, Ausems MG, Barone R, Herzog A, Heutink P, Homfray T, Gillessen-Kaesbach G, Koenig R, Kunze J, Meinecke P, Muller D, Rizzo R, Streng S, Superti-Furga A, Grzeschik KH (1999) Point mutations throughout the *GLI3* gene cause Greig cephalopolysyndactyly syndrome. *Hum Mol Genet* 8:1769-1777
- Kang S, Graham JM, Jr., Olney AH, Biesecker LG (1997) *GLI3* frameshift mutations cause autosomal dominant Pallister-Hall syndrome. *Nat Genet* 15:266-268
- Karlstrom RO, Tyurina OV, Kawakami A, Nishioka N, Talbot WS, Sasaki H, Schier AF (2003) Genetic analysis of zebrafish *gli1* and *gli2* reveals divergent requirements for gli genes in vertebrate development. *Development* 130:1549-1564
- Karpen HE, Bukowski JT, Hughes T, Gratton JP, Sessa WC, Gailani MR (2001) The sonic hedgehog receptor patched associates with caveolin-1 in cholesterol-rich microdomains of the plasma membrane. *J Biol Chem* 276:19503-19511
- Kas K, Wlodarska I, Meyen E, Van den Berghe H, Van de Ven WJ (1996) Assignment of the gene encoding human Kruppel-related zinc finger protein 4 (*GLI4*) to 8q24.3 by fluorescent in situ hybridization. *Cytogenet Cell Genet* 72:297-298

- Katoh Y, Katoh M (2004) KIF27 is one of orthologs for *Drosophila* Costal-2. *Int J Oncol* 25:1875-1880
- Kinzler KW, Bigner SH, Bigner DD, Trent JM, Law ML, O'Brien SJ, Wong AJ, Vogelstein B (1987) Identification of an amplified, highly expressed gene in a human glioma. *Science* 236:70-73
- Kinzler KW, Ruppert JM, Bigner SH, Vogelstein B (1988) The GLI gene is a member of the Kruppel family of zinc finger proteins. *Nature* 332:371-374
- Kinzler KW, Vogelstein B (1990) The GLI gene encodes a nuclear protein which binds specific sequences in the human genome. *Mol Cell Biol* 10:634-642
- Kleinjan DA, van Heyningen V (2005) Long-range control of gene expression: emerging mechanisms and disruption in disease. *Am J Hum Genet* 76:8-32
- Kogerman P, Grimm T, Kogerman L, Krause D, Unden AB, Sandstedt B, Toftgard R, Zaphiropoulos PG (1999) Mammalian suppressor-of-fused modulates nuclear-cytoplasmic shuttling of Gli-1. *Nat Cell Biol* 1:312-319
- Krauss S, Concordet JP, Ingham PW (1993) A functionally conserved homolog of the *Drosophila* segment polarity gene *hh* is expressed in tissues with polarizing activity in zebrafish embryos. *Cell* 75:1431-1444
- Kuijper S, Beverdam A, Kroon C, Brouwer A, Candille S, Barsh G, Meijlink F (2005) Genetics of shoulder girdle formation: roles of *Tbx15* and *aristaless*-like genes. *Development* 132:1601-1610
- Lalioti MD, Scott HS, Buresi C, Rossier C, Bottani A, Morris MA, Malafosse A, Antonarakis SE (1997) Dodecamer repeat expansion in cystatin B gene in progressive myoclonus epilepsy. *Nature* 386:847-851
- Larsen F, Gundersen G, Lopez R, Prydz H (1992) CpG islands as gene markers in the human genome. *Genomics* 13:1095-1107
- Lawn RM, Boonmark NW, Schwartz K, Lindahl GE, Wade DP, Byrne CD, Fong KJ, Meer K, Pathy L (1995) The recurring evolution of lipoprotein(a). Insights from cloning of hedgehog apolipoprotein(a). *J Biol Chem* 270:24004-24009
- Lettice LA, Heaney SJ, Purdie LA, Li L, de Beer P, Oostra BA, Goode D, Elgar G, Hill RE, de Graaff E (2003) A long-range *Shh* enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet* 12:1725-1735
- Lettice LA, Horikoshi T, Heaney SJ, van Baren MJ, van der Linde HC, Breedveld GJ, Joosse M, Akarsu N, Oostra BA, Endo N, Shibata M, Suzuki M, Takahashi E, Shinka T, Nakahori Y, Ayusawa D, Nakabayashi K, Scherer SW, Heutink P, Hill RE, Noji S (2002) Disruption of a long-range cis-acting regulator for *Shh* causes preaxial polydactyly. *Proc Natl Acad Sci U S A* 99:7548-7553
- Lin CY, Chen YH, Lee HC, Tsai HJ (2004) Novel cis-element in intron 1 represses somite expression of zebrafish *myf-5*. *Gene* 334:63-72
- Litingtung Y, Dahn RD, Li Y, Fallon JF, Chiang C (2002) *Shh* and *Gli3* are dispensable for limb skeleton formation but regulate digit number and identity. *Nature* 418:979-983
- Liu CZ, Yang JT, Yoon JW, Villavicencio E, Pfendler K, Walterhouse D, Iannaccone P (1998) Characterization of the promoter region and genomic organization of *GLI*, a member of the Sonic hedgehog-Patched signaling pathway. *Gene* 209:1-11
- Liu T, Wu J, He F (2000) Evolution of cis-acting elements in 5' flanking regions of vertebrate actin genes. *J Mol Evol* 50:22-30

- Manzanares M, Cordes S, Ariza-McNaughton L, Sadl V, Maruthainar K, Barsh G, Krumlauf R (1999) Conserved and distinct roles of kreisler in regulation of the paralogous Hoxa3 and Hoxb3 genes. *Development* 126:759-769
- Marigo V, Johnson RL, Vortkamp A, Tabin CJ (1996) Sonic hedgehog differentially regulates expression of GLI and GLI3 during limb development. *Dev Biol* 180:273-283
- Marigo V, Roberts DJ, Lee SM, Tsukurov O, Levi T, Gastier JM, Epstein DJ, Gilbert DJ, Copeland NG, Seidman CE, et al. (1995) Cloning, expression, and chromosomal location of SHH and IHH: two human homologues of the *Drosophila* segment polarity gene hedgehog. *Genomics* 28:44-51
- Marti E, Bovolenta P (2002) Sonic hedgehog in CNS development: one signal, multiple outputs. *Trends Neurosci* 25:89-96
- Masuda ES, Naito Y, Tokumitsu H, Campbell D, Saito F, Hannum C, Arai K, Arai N (1995) NFATx, a novel member of the nuclear factor of activated T cells family that is expressed predominantly in the thymus. *Mol Cell Biol* 15:2697-2706
- Matsumoto N, Fujimoto M, Kato R, Niikawa N (1996) Assignment of the human GLI2 gene to 2q14 by fluorescence in situ hybridization. *Genomics* 36:220-221
- Matzuk MM, Kumar TR, Vassalli A, Bickenbach JR, Roop DR, Jaenisch R, Bradley A (1995) Functional analysis of activins during mammalian development. *Nature* 374:354-356
- McCaffrey PG, Luo C, Kerppola TK, Jain J, Badalian TM, Ho AM, Burgeon E, Lane WS, Lambert JN, Curran T, et al. (1993) Isolation of the cyclosporin-sensitive T cell transcription factor NFATp. *Science* 262:750-754
- McCarthy RA, Barth JL, Chintalapudi MR, Knaak C, Argraves WS (2002) Megalin functions as an endocytic sonic hedgehog receptor. *J Biol Chem* 277:25660-25667
- McDermott A, Gustafsson M, Elsam T, Hui CC, Emerson CP, Jr., Borycki AG (2005) Gli2 and Gli3 have redundant and context-dependent function in skeletal muscle formation. *Development* 132:345-357
- McMahon AP (2000) More surprises in the Hedgehog signaling pathway. *Cell* 100:185-188
- Merchant M, Vajdos FF, Ultsch M, Maun HR, Wendt U, Cannon J, Desmarais W, Lazarus RA, de Vos AM, de Sauvage FJ (2004) Suppressor of fused regulates Gli activity through a dual binding mechanism. *Mol Cell Biol* 24:8627-8641
- Methot N, Basler K (2000) Suppressor of fused opposes hedgehog signal transduction by impeding nuclear accumulation of the activator form of Cubitus interruptus. *Development* 127:4001-4010
- Miyatake S, Shlomai J, Arai K, Arai N (1991) Characterization of the mouse granulocyte-macrophage colony-stimulating factor (GM-CSF) gene promoter: nuclear factors that interact with an element shared by three lymphokine genes--those for GM-CSF, interleukin-4 (IL-4), and IL-5. *Mol Cell Biol* 11:5894-5901
- Mo R, Freer AM, Zinyk DL, Crackower MA, Michaud J, Heng HH, Chik KW, Shi XM, Tsui LC, Cheng SH, Joyner AL, Hui C (1997) Specific and redundant functions of Gli2 and Gli3 zinc finger genes in skeletal patterning and development. *Development* 124:113-123

- Monnier V, Dussillol F, Alves G, Lamour-Isnard C, Plessis A (1998) Suppressor of fused links fused and Cubitus interruptus on the hedgehog signalling pathway. *Curr Biol* 8:583-586
- Morrison A, Ariza-McNaughton L, Gould A, Featherstone M, Krumlauf R (1997) HOXD4 and regulation of the group 4 paralog genes. *Development* 124:3135-3146
- Motoyama J, Liu J, Mo R, Ding Q, Post M, Hui CC (1998) Essential function of Gli2 and Gli3 in the formation of lung, trachea and oesophagus. *Nat Genet* 20:54-57
- Motoyama J, Milenkovic L, Iwama M, Shikata Y, Scott MP, Hui CC (2003) Differential requirement for Gli2 and Gli3 in ventral neural cell fate specification. *Dev Biol* 259:150-161
- Niemann H, Kues WA (2000) Transgenic livestock: premises and promises. *Anim Reprod Sci* 60-61:277-293
- Nobrega MA, Ovcharenko I, Afzal V, Rubin EM (2003) Scanning human gene deserts for long-range enhancers. *Science* 302:413
- Nobrega MA, Pennacchio LA (2004) Comparative genomic analysis as a tool for biological discovery. *J Physiol* 554:31-39
- Paces-Fessy M, Boucher D, Petit E, Paute-Briand S, Blanchet-Tournier MF (2004) The negative regulator of Gli, Suppressor of fused (Sufu), interacts with SAP18, Galectin3 and other nuclear proteins. *Biochem J* 378:353-362
- Park HL, Bai C, Platt KA, Matisse MP, Beeghly A, Hui CC, Nakashima M, Joyner AL (2000) Mouse Gli1 mutants are viable but have defects in SHH signaling in combination with a Gli2 mutation. *Development* 127:1593-1605
- Pavletich NP, Pabo CO (1993) Crystal structure of a five-finger GLI-DNA complex: new perspectives on zinc fingers. *Science* 261:1701-1707
- Pepinsky RB, Zeng C, Wen D, Rayhorn P, Baker DP, Williams KP, Bixler SA, Ambrose CM, Garber EA, Miatkowski K, Taylor FR, Wang EA, Galdes A (1998) Identification of a palmitic acid-modified form of human Sonic hedgehog. *J Biol Chem* 273:14037-14045
- Perrimon N (1995) Hedgehog and beyond. *Cell* 80:517-520
- Rahnama F, Toftgard R, Zaphiropoulos PG (2004) Distinct roles of PTCH2 splice variants in Hedgehog signalling. *Biochem J* 378:325-334
- Rallis C, Del Buono J, Logan MP (2005) Tbx3 can alter limb position along the rostrocaudal axis of the developing embryo. *Development* 132:1961-1970
- Rallu M, Corbin JG, Fishell G (2002) Parsing the prosencephalon. *Nat Rev Neurosci* 3:943-951
- Riddle RD, Johnson RL, Laufer E, Tabin C (1993) Sonic hedgehog mediates the polarizing activity of the ZPA. *Cell* 75:1401-1416
- Ristevski S (2005) Making better transgenic models: conditional, temporal, and spatial approaches. *Mol Biotechnol* 29:153-163
- Robbins DJ, Nybakken KE, Kobayashi R, Sisson JC, Bishop JM, Therond PP (1997) Hedgehog elicits signal transduction by means of a large complex containing the kinesin-related protein costal2. *Cell* 90:225-234
- Roberts WM, Douglass EC, Peiper SC, Houghton PJ, Look AT (1989) Amplification of the gli gene in childhood sarcomas. *Cancer Res* 49:5407-5413
- Roelink H, Porter JA, Chiang C, Tanabe Y, Chang DT, Beachy PA, Jessell TM (1995) Floor plate and motor neuron induction by different concentrations of the amino-terminal cleavage product of sonic hedgehog autoproteolysis. *Cell* 81:445-455

- Roessler E, Du YZ, Mullor JL, Casas E, Allen WP, Gillesen-Kaesbach G, Roeder ER, Ming JE, Ruiz i Altaba A, Muenke M (2003) Loss-of-function mutations in the human GLI2 gene are associated with pituitary anomalies and holoprosencephaly-like features. *Proc Natl Acad Sci U S A* 100:13424-13429
- Roessler E, Ermilov AN, Grange DK, Wang A, Grachtchouk M, Dlugosz AA, Muenke M (2005) A previously unidentified amino-terminal domain regulates transcriptional activity of wild-type and disease-associated human GLI2. *Hum Mol Genet* 14:2181-2188
- Ruiz i Altaba A (1999) Gli proteins encode context-dependent positive and negative functions: implications for development and disease. *Development* 126:3205-3216
- Ruppert JM, Kinzler KW, Wong AJ, Bigner SH, Kao FT, Law ML, Seunavaz HN, O'Brien SJ, Vogelstein B (1988) The GLI-Kruppel family of human genes. *Mol Cell Biol* 8:3104-3113
- Ruppert JM, Vogelstein B, Arheden K, Kinzler KW (1990) GLI3 encodes a 190-kilodalton protein with multiple regions of GLI similarity. *Mol Cell Biol* 10:5408-5415
- Sagai T, Hosoya M, Mizushima Y, Tamura M, Shiroishi T (2005) Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. *Development* 132:797-803
- Salinas PC (2003) The morphogen sonic hedgehog collaborates with netrin-1 to guide axons in the spinal cord. *Trends Neurosci* 26:641-643
- Sasaki H, Nishizaki Y, Hui C, Nakafuku M, Kondoh H (1999) Regulation of Gli2 and Gli3 activities by an amino-terminal repression domain: implication of Gli2 and Gli3 as primary mediators of Shh signaling. *Development* 126:3915-3924
- Schimmang T, van der Hoeven F, Ruether U (1993) Gli3 expression is affected in the morphogenetic mouse mutants add and Xt. *Prog Clin Biol Res* 383A:153-161
- Schwartz C, Locke J, Nishida C, Kornberg TB (1995) Analysis of cubitus interruptus regulation in *Drosophila* embryos and imaginal disks. *Development* 121:1625-1635
- Schwartz S, Zhang Z, Frazer KA, Smit A, Riemer C, Bouck J, Gibbs R, Hardison R, Miller W (2000) PipMaker--a web server for aligning two genomic DNA sequences. *Genome Res* 10:577-586
- Schweitzer R, Vogan KJ, Tabin CJ (2000) Similar expression and regulation of Gli2 and Gli3 in the chick limb bud. *Mech Dev* 98:171-174
- Sekimizu K, Nishioka N, Sasaki H, Takeda H, Karlstrom RO, Kawakami A (2004) The zebrafish iguana locus encodes Dzip1, a novel zinc-finger protein required for proper regulation of Hedgehog signaling. *Development* 131:2521-2532
- Simmons AD, Horton S, Abney AL, Johnson JE (2001) Neurogenin2 expression in ventral and dorsal spinal neural tube progenitor cells is regulated by distinct enhancers. *Dev Biol* 229:327-339
- Sisson JC, Ho KS, Suyama K, Scott MP (1997) Costal2, a novel kinesin-related protein in the Hedgehog signaling pathway. *Cell* 90:235-245
- St-Jacques B, Hammerschmidt M, McMahon AP (1999) Indian hedgehog signaling regulates proliferation and differentiation of chondrocytes and is essential for bone formation. *Genes Dev* 13:2072-2086
- Sumarsono SH, Wilson TJ, Tymms MJ, Venter DJ, Corrick CM, Kola R, Lahoud MH, Papas TS, Seth A, Kola I (1996) Down's syndrome-like skeletal abnormalities in *Ets2* transgenic mice. *Nature* 379:534-537

- Tanimura A, Dan S, Yoshida M (1998) Cloning of novel isoforms of the human Gli2 oncogene and their activities to enhance tax-dependent transcription of the human T-cell leukemia virus type 1 genome. *J Virol* 72:3958-3964
- Tay SY, Ingham PW, Roy S (2005) A homologue of the Drosophila kinesin-like protein Costal2 regulates Hedgehog signal transduction in the vertebrate embryo. *Development* 132:625-634
- Topp J (1997) Funktionelle Charakterisierung des Kandidatengens für das Greig Cephalopolysyndaktylie Syndrom, GLI3. Diplomwork, Philipps-University Marburg, Germany
- Tyurina OV, Guner B, Popova E, Feng J, Schier AF, Kohtz JD, Karlstrom RO (2005) Zebrafish Gli3 functions as both an activator and a repressor in Hedgehog signaling. *Dev Biol* 277:537-556
- van der Hoeven F, Schimmang T, Vortkamp A, Ruether U (1993) Molecular linkage of the morphogenetic mutation add and the zinc finger gene Gli3. *Mamm Genome* 4:276-277
- Venkatesh B, Yap WH (2005) Comparative genomics using fugu: a tool for the identification of conserved vertebrate cis-regulatory elements. *Bioessays* 27:100-107
- Villavicencio EH, Walterhouse DO, Iannaccone PM (2000) The sonic hedgehog-patched-gli pathway in human development and disease. *Am J Hum Genet* 67:1047-1054
- Vortkamp A, Gessler M, Grzeschik KH (1991) GLI3 zinc-finger gene interrupted by translocations in Greig syndrome families. *Nature* 352:539-540
- Vortkamp A, Gessler M, Grzeschik KH (1995) Identification of optimized target sequences for the GLI3 zinc finger protein. *DNA Cell Biol* 14:629-634
- Vortkamp A, Heid C, Gessler M, Grzeschik KH (1995) Isolation and characterization of a cosmid contig for the GCPS gene region. *Hum Genet* 95:82-88
- Vortkamp A, Lee K, Lanske B, Segre GV, Kronenberg HM, Tabin CJ (1996) Regulation of rate of cartilage differentiation by Indian hedgehog and PTH-related protein. *Science* 273:613-622
- Walters M, Martin DI (1992) Functional erythroid promoters created by interaction of the transcription factor GATA-1 with CACCC and AP-1/NFE-2 elements. *Proc Natl Acad Sci U S A* 89:10444-10448
- Walther N, Jansen M, Ergun S, Kascheike B, Ivell R (1996) Sertoli cell lines established from H-2Kb-tsA58 transgenic mice differentially regulate the expression of cell-specific genes. *Exp Cell Res* 225:411-421
- Wang B, Fallon JF, Beachy PA (2000) Hedgehog-regulated processing of Gli3 produces an anterior/posterior repressor gradient in the developing vertebrate limb. *Cell* 100:423-434
- Wang QT, Holmgren RA (1999) The subcellular localization and activity of Drosophila cubitus interruptus are regulated at multiple levels. *Development* 126:5097-5106
- Warburton D, Lee MK (1999) Current concepts on lung development. *Curr Opin Pediatr* 11:188-192
- Werner CA, Dohner H, Joos S, Trumper LH, Baudis M, Barth TF, Ott G, Moller P, Lichter P, Bentz M (1997) High-level DNA amplifications are common genetic aberrations in B-cell neoplasms. *Am J Pathol* 151:335-342
- Whitelaw E, Tsai SF, Hogben P, Orkin SH (1990) Regulated expression of globin chains and the erythroid transcription factor GATA-1 during erythropoiesis in the developing mouse. *Mol Cell Biol* 10:6596-6606

- Wild A, Kalff-Suske M, Vortkamp A, Bornholdt D, Koenig R, Grzeschik KH (1997) Point mutations in human GLI3 cause Greig syndrome. *Hum Mol Genet* 6:1979-1984
- Wolff C, Roy S, Lewis KE, Schauerte H, Joerg-Rauch G, Kirn A, Weiler C, Geisler R, Haffter P, Ingham PW (2004) *iguana* encodes a novel zinc-finger protein with coiled-coil domains essential for Hedgehog signal transduction in the zebrafish embryo. *Genes Dev* 18:1565-1576
- Wolvetang EJ, Wilson TJ, Sanij E, Busciglio J, Hatzistavrou T, Seth A, Hertzog PJ, Kola I (2003) ETS2 overexpression in transgenic models and in Down syndrome predisposes to apoptosis via the p53 pathway. *Hum Mol Genet* 12:247-255
- Xie X, Lu J, Kulbokas EJ, Golub TR, Mootha V, Lindblad-Toh K, Lander ES, Kellis M (2005) Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature* 434:338-345
- Yang JT, Liu CZ, Villavicencio EH, Yoon JW, Walterhouse D, Iannaccone PM (1997) Expression of human GLI in mice results in failure to thrive, early death, and patchy Hirschsprung-like gastrointestinal dilatation. *Mol Med* 3:826-835
- Zoltewicz JS, Stewart NJ, Leung R, Peterson AS (2004) Atrophin 2 recruits histone deacetylase and is required for the function of multiple signaling centers during mouse embryogenesis. *Development* 131:3-14

10 PUBLICATIONS

10.1 Original work

Kalff-Suske M, Paparidis Z, Bornholdt D, Cole T, Kalff-Suske M, Grzeschik KH (2004) Gene symbol: GLI3. Disease: Pallister-Hall syndrome. Hum Genet 114:403

10.2 Posters

Paparidis, Z., Jacobsen, E., Kalff-Suske, M., Grzeschik, K.-H. (2004) Transcriptional control of GLI3 gene expression. European Human Genetics Conference 2004 in München/Germany (June 12-15, 2004)

Paparidis, Z., de Graaff E., Elgar, G., Grzeschik, K.-H. (2005) Transcriptional control of GLI gene expression. 16th Annual Meeting of the German Society of Human Genetics in Halle/Germany (March 9-12, 2005)

Paparidis Z., de Graaff E., Elgar G., Grzeschik K.-H- (2005) Highly conserved noncoding DNA sequences within introns are controlling *in cis* the expression of GLI3. European Human Genetics Conference 2005 in Prague/Czech Republic (May 7-10, 2005)

10.3 Seminars

Functional analysis of genes (*GLI*) involved in limb development. Progress report at the meeting of the International Graduiertenkolleg GRK767, 9th Nov. 2002, IMT Marburg, Germany

Functional analysis of genes (*GLI*) involved in limb development. 1st Winter School of the International Graduiertenkolleg “Transcriptional Control in Developmental Processes” Marburg-Giessen-Rotterdam at the Kleinwalsertal, 15-20 February 2003, Kleinwalsertal, Austria

Functional analysis of the GLI genes involved in limb development. 2nd Winter School of the International Graduiertenkolleg “Transcriptional Control in Developmental Processes” Marburg-Giessen-Rotterdam at the Kleinwalsertal, 13-18 March 2004, Kleinwalsertal, Austria

11 ACKNOWLEDGEMENTS

First of all I would like to thank Prof. Dr. Renate Renkawitz-Pohl for the supervision of my project and for the opportunity to aim for my promotion in Biology. A big thanks goes to Prof. Dr. Karl-Heinz Grzeschik for providing me with the subject for my thesis. Without his constant help, advice, and supervision the accomplishment of this work would not have been possible.

I am also grateful to Dr. Esther deGraaff from the Department of Clinical Genetics of the Erasmus MC in Rotterdam for giving me the opportunity to work for one month in her group, to learn new techniques, and for helping me with the transgenic mice. Valuable clinical advice by Prof. Dr. Manuela C. Koch is gratefully acknowledged.

Of course I wish to thank Kester Dürthi, Dr. Stefan Zauner and Oliver Kawach of the Biological Faculty of the Philipps University Marburg for their essential help with the primer extension experiments and Jörg Leers of the Genetics Institute of the Justus-Liebig University of Giessen for his support with DNA-protein bandshift assays.

I am greatly thankful to Prof. Dr. Jürgen Kunz and Ulrike Neidel for supporting me with the sequencing experiments. In addition, credit must be given to my other colleagues at the Institut of Human Genetics, Philipps University Marburg, Germany, for their collaboration and for creating a pleasant environment: Sajid Malik, Amir Abbasi, Claudia Moch, Shalini Singh, Dorothea Bornholdt (thanks for the many nice suggestions), Anja Volz-Peters, Janna Riemann, Dr. Leonora Leveleki, Mark Rauschendorf and Gerti Panzner. Also, Steffen Uebe and Harald Balz have been a great help with the computers. An extra thanks to our secretary Silke Janitz for being so polite and helpful anytime. In addition I am grateful to Dr. Martha Kalff Suske for her help and suggestions during my work

I like to thank the Deutsche Forschungsgemeinschaft and the Graduiertenkolleg (GRK 767) for funding my research project as well as the opportunities for attending scientific conferences and collaborating with other working groups.

Finally, I would like to express my gratitude towards my parents and my family for being always there when I needed them and for providing me with the strength to reach the level at which I am now. This work is dedicated to them with love. I would also like to dedicate it to Alexandros Argyrokastritis, a very special person and teacher that unfortunately left us too soon.

12 DECLARATION

I hereby declare that I have worked on my thesis “Cis-acting elements controlling the expression of the human GLI3 gene” independently, without unauthorized assistance and that I did not use any other than my clearly stated sources and help.

13 CURRICULUM VITAE

Personal details

Name: Paparidis Zissis
Date of birth: 27th June 1979
Marital status: Single
Nationality: Greek

Address:

Work: Center for Human Genetics
Philipps-University Marburg
Bahnhofstr. 7
35037 Marburg
Germany

Home: Fuchspass 32
35039
Marburg
Germany
Tel: 01788187352
E-mail: zissman@hotmail.com

Education and Qualifications

September 2002-
September 2005

**Institute of Human Genetics, Philipps University
Marburg**

PhD Thesis: ‘Cis-acting elements controlling the expression of the human GLI3 gene’
During this period I have learned molecular techniques such as primer extension, 5’ RACE PCR, cloning (ligation, transformation, DNA extraction, restriction enzyme analysis, agarose gel electrophoresis), sequencing, point mutagenesis, cell transfection and dual luciferase assay. I was also trained in mutation analysis through SSCA, in cancer cell and stem cell culturing. Finally, I familiarized myself with DNA-protein band shift assay techniques.

September 2001 –
June 2002

**Biochemistry lab of the University of Crete,
Department of Biology, Faculty of Natural and
Technological Sciences**

Diploma Thesis: ‘Study of the lipoprotein gene of the terrestrial crab *Potamon potamios* (brachyura)’ under the supervision of professor Stratakis E.

September 2000 –
July 2001

**Biochemistry laboratory - University of Crete,
Department of Biology, Faculty of Natural and
Technological Sciences**

Isolation and biochemical identification of proteins (lectins from marine invertebrates). Molecular cloning of lectin and actin genes from marine invertebrates. During this period I have learned biochemical techniques (affinity chromatography, agglutination test, SDS and native polyacrylamide electrophoresis). I also familiarized myself with molecular techniques such as RNA extraction, RT-PCR (Reverse Transcription-Polymerase Chain Reaction), PCR, TA cloning (ligation, transformation, DNA extraction, restriction enzyme analysis, agarose gel electrophoresis), nucleic acid hybridization. (This work was carried out at the Institute of Molecular Biology and Biotechnology in Heraklion).

April – June 1999

**Marine Biology laboratory – University of Crete,
Department of Biology, Faculty of Natural and
Technological Sciences**

First contact with molecular techniques.

1998 – 2002

**University of Crete, Department of Biology, Faculty
of Natural and Technological Sciences**

Final grade: 7.65

Courses taken: Microbiology, Physics, Organic Chemistry I – II, Zoology I - II, English I - IV, Computing, Bio-Mathematics, Plant Biology I – II, Classical Genetics, Molecular Genetics, Biochemistry I – II, Molecular Biology, Ecology, General Chemistry, Cell Biology, Biotechnology I - II, Human genetics, Immunobiology, Evolutionary Biology, Introduction to Medical Entomology, Molecular Oncogenesis, Modern Microscopy, Applied Biology, German I - IV, Regulation of gene – expression, Molecular physiology.

Gained experience by attending laboratory exercises.

Field course experience

Animal and plant sampling and species identification. Visits to cretan mountains, fields and beaches in order to examine various ecosystems and spot their differences and aiming to study the effects of different amounts of pollution.

1997 – 1998

Exams for University admission

Biology (18.2/20), Chemistry (18.1/20), Physics (16.2/20),

1991 – 1997

German School of Thessaloniki

Final grade: Very good (16 7/11)

Anthropology – Biology (17/20), Physics (19/20),
Mathematics (18/20), Chemistry (17/20), Grammar,
History, Philosophy, English, German.